Technical Report 1047

Hypothesizing Device Mechanisms: Opening Up the Black Box

Richard James Doyle

MIT Artificial Intelligence Laboratory

This blank page was inserted to preserve pagination.

Hypothesizing Device Mechanisms: Opening Up the Black Box

by

Richard James Doyle

June 1988

This report is a revised version of a thesis submitted on May 5, 1988 to the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research Contract N00014-85-K-0124.

Hypothesizing Device Mechanisms: Opening Up the Black Box

by
Richard James Doyle

Submitted to the Department of Electrical Engineering and Computer Science
on May 5, 1988, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract: Causal models of devices support many forms of problem solving in the physical system domain, such as diagnosis and monitoring. I describe an approach to forming hypotheses about hidden mechanism configurations within devices given external observations and a vocabulary of primitive mechanisms. The approach has two aspects: one involves a set of constraints drawn from physical and causal principles to prune hypotheses; the other involves an ordering on hypothesis types and a set of rules for traversing the ordering to carefully control the generation of hypotheses. The rules are all based on the principle that incomplete hypotheses exhibit characteristic deficiencies; they justify attempts to augment deficient hypotheses by extending them into more complex hypotheses.

This approach has been implemented in a causal modelling system called JACK. The program JACK generates manageably sized sets of hypotheses about the mechanisms within devices and makes fine distinctions among hypotheses. This causal modelling system reasons about the behavior of several diverse devices, constructing explanations for why a second piece of toast in a toaster comes out lighter, why the slide in a tire gauge does not slip back inside the cylinder when the gauge is removed from the tire, and how in a refrigerator a single substance can serve alternately as a heat sink for the interior and a heat source for the exterior.

I analyze the performance of the program JACK in two ways: in terms of the number of hypotheses admitted for each device example and how these hypotheses are organized in an abstraction space, and in terms of empirical results from a set of experiments which isolate the pruning power due to the different sources of constraint in my approach to the causal modelling problem. In conclusion, I show how causal models of devices produced by the program JACK can be used to support diagnosis and monitoring tasks.

Thesis Supervisor: Patrick H. Winston
Title: Professor of Computer Science

Dedicated

to my parents,
who share deeply in my accomplishment.

Acknowledgements

Here I make a humble attempt to recall those who played a part in this project, and in my life during it. I thank:

Patrick Winston, my thesis supervisor, for lucid advice and for gentle encouragement when I needed it.

Randy Davis and Tomàs Lozano-Pèrez, my thesis readers, for keeping track of and shaping this thesis.

Mike Kashket, for collation, humidity, choisteness, and other things even more obscure. Muchly.

David Kirsh, for being a solid friend and for the charm of his uniquely contemplative yet earthy view of life.

Boris Katz, for looking after me in a fatherly way.

Jonathan Amsterdam and Karl Ulrich, who read this thesis cover to cover, and the other members of Patrick Winston's research group.

My officemates over the years, Michael Brent, Ken Yip, and Bob Givan, for making room 823 a special place within the AI Lab.

My colleagues and friends at JPL, Dave Atkinson, Suzanne Sellers, Raj Doshi, Harry Porta, and the other members of the AI group.

Hake and Gary, my two oldest friends in the world, for affection and laughter.

Cecilia Guiar, for love and friendship from afar through all these years.

And to persons and things purple.

TABLE OF CONTENTS

1. Causal Modelling: Figuring Out How Things Work

The process of constructing and refining physical models to account for observations is a fair characterization of what science is all about. In this work, I investigate the modelling process itself. The domain is devices, or designed physical systems. The research goal is to articulate a set of principles which engender capabilities for hypothesizing manageably small sets of physically plausible device models, and for making fine distinctions among those models.

The modelling problem for devices may be posed in several ways: "What hidden configuration of mechanisms can explain this behavior?", "Is this hypothesis consistent with all observations?", "How may have this device been designed?". I have developed a modelling system—called JACK—which addresses these questions and produces abstract causal models of several physical systems, including a toaster, a pocket tire gauge, a bicycle drive, a refrigerator, and a home heating system.

The importance of the modelling problem arises from its ubiquity. The need to understand how things work inevitably arises in the course of other problem solving tasks. In the physical system domain these other problem solving tasks include diagnosis, monitoring, and planning. For example, identifying a fault in a device requires knowing how it was supposed to work in the first place and hypothesizing how a fault explains the observed behavior.

The modelling problem, or "black box" problem, is notoriously difficult. The difficulty is traceable in part to the great number of potential hypotheses. The number of possible hypotheses for even a relatively simple device like a toaster, within the representations used by the program JACK, is in the several millions.

My approach to making the modelling problem tractable in the physical system domain is two-pronged. One of the prongs involves applying a set of constraints which embody physical and causal principles to prune hypotheses. The other prong involves enumerating different forms for hypotheses, placing an ordering on these forms, and using this ordering to carefully control the generation of hypotheses. The pruning power resulting from the combined application of these two thrusts has proven to be impressive. The program JACK generates on the order of a hundred hypotheses for the toaster. Among these is an abstraction of the standard design for toasters.

1.1 Scenarios

Before offering a set of scenarios which delineate specific performance goals

for a causal modelling system, I must state the following caveat—my work is not cognitive science. Although I have been inspired by attempts on the part of people at modelling devices, including my own, I have no goal of gaining insight into human performance at modelling devices and make no claims along these lines. The scenarios developed in the next few paragraphs have two purposes: First, to provide the reader with a greater grasp of the causal modelling task and its difficulties in advance of the forthcoming details, and second, to provide benchmarks against which to evaluate the spirit, if not the absolute letter, of the reasoning exhibited by the program JACK.

1.1.1 A Toaster

The single most mysterious aspect of a toaster is the nature of the mechanism whereby bread stays down in the toaster for an apparently measured amount of time. Most people have no trouble conjecturing that the downward motion of the lever closes a switch, that the coils are electrically heated, that the bread turns dark because it is heated, and that the carriage is spring-loaded. However, the timing mechanism within a toaster which is responsible for producing toast of just such a darkness is more puzzling. Curiously, it is a common form of toaster *mis*behavior which provides a clue to the nature of this mechanism.

People typically offer two hypotheses concerning the timing device within a toaster: one like an alarm clock, the other like a thermostat. The alarm clock hypothesis involves a motor powered by electricity which steadily moves a latch on a spring until it disengages—at this point the toast pops up. The thermostat hypothesis also involves a moving latch on a spring, but here the cause of the motion is thermal expansion due to heating within the toaster, not an electric motor.

Both of these hypotheses can explain a single example of toaster operation. However, only one can explain the annoying and familiar behavior where a second piece of bread placed in a toaster shortly after a first piece turns out lighter. The causal explanation for this misbehavior is that the second timing episode begins at a higher initial temperature. The already partially expanded latch has a shorter distance to expand through before the spring is released. The bread is heated for a shorter time and comes out lighter.

In the alarm clock model for a toaster, the initial temperature has no effect on the length of the period during which the bread is in the toaster. The bread is heated for the same amount of time in both episodes and should turn out darker owing to the higher initial temperature, certainly not lighter.

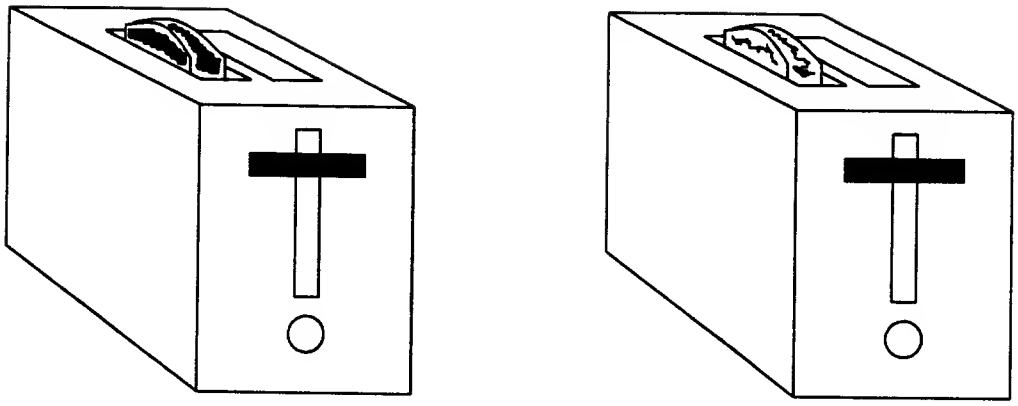


Figure 1.1. Why does a second piece of toast turn out lighter?

The causal modelling system—without knowing that the behavior involving the lighter second piece of toast falls outside the intended functioning of a toaster—is able to reproduce the reasoning outlined above and utilize this instance of toaster misbehavior to distinguish the alarm clock and thermostat hypotheses.

1.1.2 A Pocket Tire Gauge

The pocket tire gauge is an excellent example of a device for which the modelling problem is surprisingly thorny. Its range of behavior is quite small, yet this behavior is baffling. I have posed this problem to several people and few have been able to solve it.

No one ever has any trouble conjecturing that the motion of the slide in a tire gauge is a response to air pressure. But why doesn't the slide slam all the way to the end of the cylinder? One possible explanation involves an equilibrium state within the cylinder. There may be an opposing force—due to a spring, for example—which balances the air pressure. However, why doesn't the slide slip back into the cylinder when the gauge is removed from the tire? The conjectured spring force then should be the only active one.

At this point, most people become stumped. To get past this quandary, one has to note that there are couplings which allow motion in one direction but not in the opposite direction. One of these is a ratchet. However, once

again observation does not provide confirmation. The slide may be pushed easily back into the cylinder when the gauge is off the tire.

Fortunately, there is another kind of one-way coupling which is consistent with all of the observable behavior of the tire gauge. This is a coupling based simply on contact, not attachment, with which it is possible to push, but not to pull.

When the gauge is placed on a tire, released air enters the cylinder and pushes a piston inside the cylinder. This piston eventually touches and then pushes the slide. The piston is spring-loaded so that its motion is arrested when the restoring force of the spring exactly balances the force due to the air pressure. The slide, no longer being pushed by the piston, also stops moving.

When the gauge is removed from the tire, the force due to air pressure disappears and the now-unopposed spring pushes the piston back into the cylinder. However, the slide—unattached to the piston—stays right where it is.

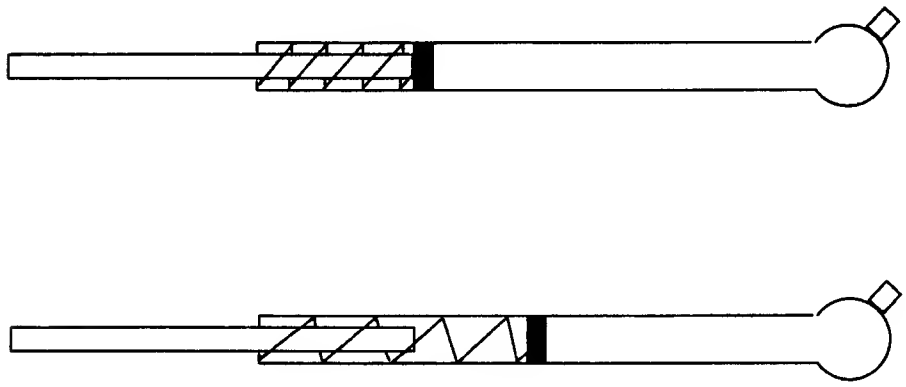


Figure 1.2. Why does the slide remain stationary?

The design of the pocket tire gauge is elegant. This design proves obscure for most people. Nevertheless, the program JACK, embodying the approach to modelling described in this thesis, is able to solve this modelling task.

1.1.3 A Refrigerator

Refrigerators are a complete riddle to most people. The physical principles on which they operate are not in the vocabulary of the typical layperson.

These principles are: (1) the boiling point of a substance is a function of the ambient pressure and (2) condensation and evaporation are, respectively, heat-releasing and heat-absorbing processes.

The interior of a refrigerator is cooled by forcing the evaporation of a substance through a sudden drop in pressure. The substance absorbs heat from the interior while evaporating. If this were the end of the story, there might be a problem, for the implication here is that this substance serves as a heat sink of arbitrary capacity. One way to solve this problem is to continually renew this substance.

Instead, this problem is better addressed through an elegant use of synergy. The forced evaporation described above is only one half of a cycle of operations inside a refrigerator. The other half involves forced condensation of the same substance, brought about by a pressure increase. During condensation, the substance gives up the heat gained during evaporation; there is no net heat gain or loss within this substance. The condensation half resets the evaporation half of the cycle, allowing more heat to be absorbed safely from the interior the next time around.

The pressure increase which forces condensation is usually achieved through the use of a mechanical compressor. However, an equally effective pressure increase can be achieved through—paradoxically – a temperature increase. For a while, a refrigerator design which employed gas heating was competitive with the compressor design. The heat exchange mechanisms in this absorption type of refrigerator are, not surprisingly, more complicated.

The program JACK is able to model the cycle of operations within a refrigerator by recognizing that condensation and evaporation imply hidden heat sources and heat sinks and reasoning that sources and sinks may be avoided by alternating gains and losses within a cycle.

1.2 The Causal Modelling Task

The task of the causal modelling system JACK is to conjecture configurations of mechanisms inside the "black box" which are consistent with the externally observable behavior of a device.

There are two inputs to the causal modelling system: one is a description of the externally observable behavior of a device; the other is a set of mechanisms. The output is a set of compositions of those mechanisms, each explaining the behavior of the device. See Figure 1.4.

The description of the behavior of a device consists of a timeline in which changes in the observable quantities of the device are recorded. For example, part of the description of the behavior of a tire gauge involves changes in the

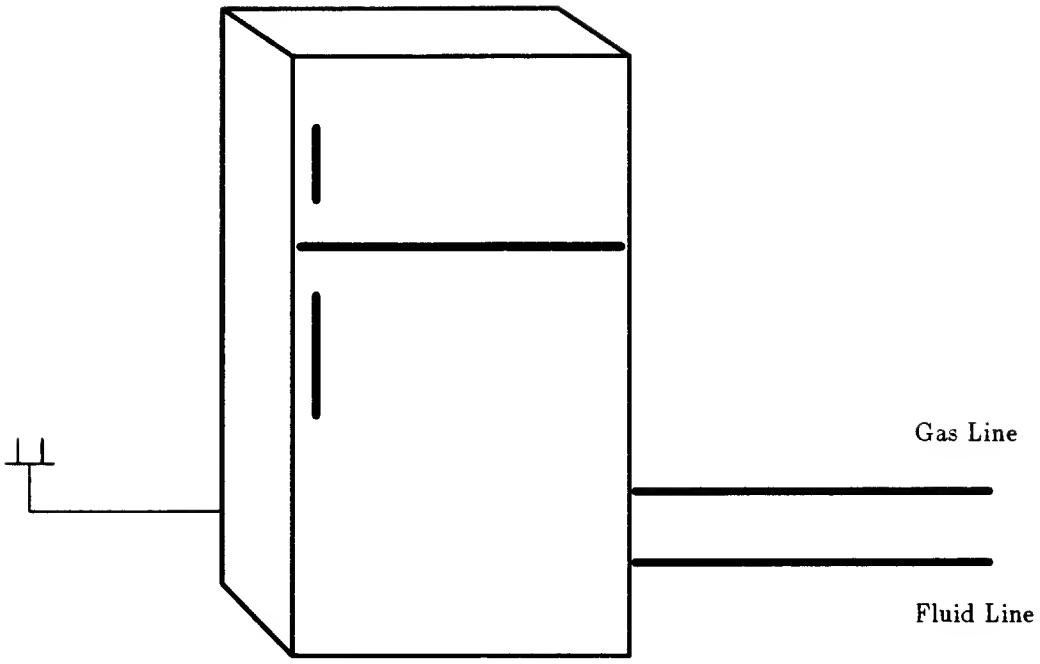


Figure 1.3. Why is this a poor design for a refrigerator?

position of the slide. Initially, the slide is stationary. Some time later, it moves out of the cylinder, reaching a new stationary position. The slide does not move again.

Examples of mechanisms are mechanical couplings, thermal expansion, fluid flow, condensation, gravity, springs, valves, etc. These mechanisms serve as the primitive causal explanations from which the model of a device is constructed. They map causes to effects. For example, a mechanical coupling maps the motion of one physical object to the motion of another physical object.

The causal modelling task is to hypothesize paths of mechanisms through the black box. These causal paths map the primitive causes of a device or its inputs, to its final effects or outputs.

Appendix A lists observations of the devices on which the program JACK has been tested. Appendix B contains the vocabulary of mechanisms used in all of the examples. Figure 1.5 shows a graphical representation of the

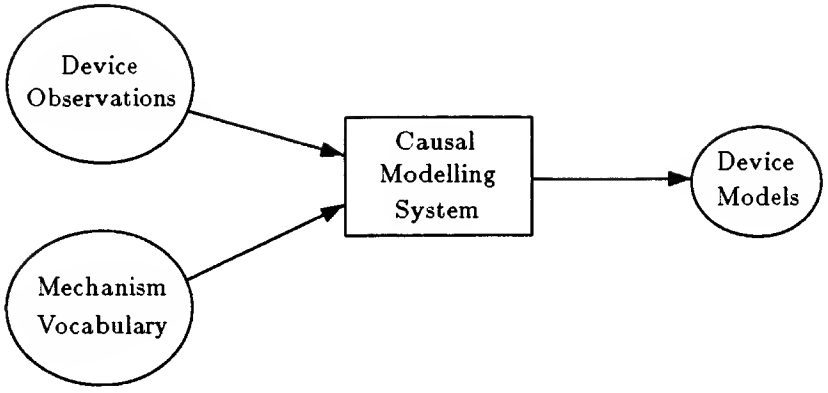


Figure 1.4. The causal modelling task.

observation of a toaster.

Figure 1.6 shows a causal graph generated by the program JACK to explain part of the observation of a toaster. The task is to explain how current at the outlet can result in heating at the coils. The graph represents the following hypothesis: Electricity flows from the outlet to the coils; at the coils current is transformed into heat. Furthermore, the motion of the lever closes a switch which enables the electrical flow.

1.3 Issues

Here, I briefly enumerate the issues which are central to this work. In the body of the thesis, I enlarge on the principles which underlie the performance of the program JACK in the context of discussing the operation of the causal modelling system and the set of device examples I have successfully implemented.

How to constrain the formation of hypotheses?

The hypothesis space for the causal modelling problem turns out to be exponential in the worst case in both the length of causal paths and in the number of interacting causal paths. Some form of strongly constrained search clearly is called for. This is the central computational issue of the thesis.

What are the constraints in the physical system domain?

On the testing end of hypothesis formation, I have devised a set of general constraints for the physical system domain from physical and causal principles.

	0:00	1:00	1:01	1:06	3:06	3:07	7:30
Lever Position Amount	Up		Down			Up	
Lever Position Rate	Zero	Negative	Zero		Positive	Zero	
Dial Angle Amount	LM						
Dial Angle Rate	Zero						
Carriage Position Amount	Up		Down			Up	
Carriage Position Rate	Zero	Negative	Zero		Positive	Zero	
Coils Temperature Amount	Off				Hot		Off
Coils Temperature Rate	Zero		Positive		Zero	Negative	Zero
Bread Appearance Amount	Untoasted				Golden		
Bread Appearance Rate	Zero			Positive	Zero		
Outlet Charge Amount	On						
Outlet Charge Rate	Positive						
Earth Gravity Amount	G						
Earth Gravity Rate	Zero						

Figure 1.5. Observation of a toaster.

The principles embodied in these constraints include conservation of energy and mass, entropy, inertia, no action at a distance, mechanical advantage, and the directionality of causation. These constraints reflect necessary conditions which all physically realized devices must satisfy.

What are the different causal structures for devices?

On the generation end of hypothesis formation, I have enumerated a set of hypothesis forms corresponding to different causal structures for devices. These structures include simple linear mechanism chains from inputs to outputs, enablement, disablement, and equilibrium interactions where the causal

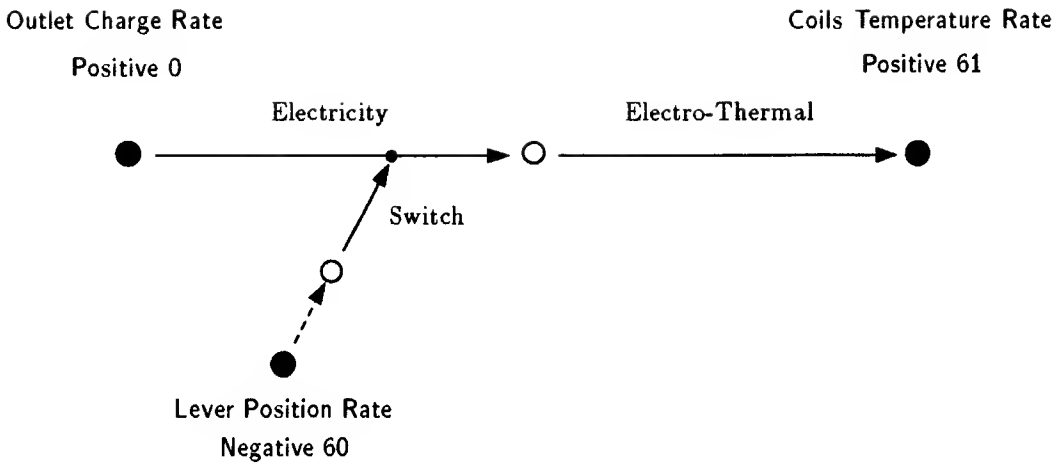


Figure 1.6. A causal graph.

contributions of different mechanisms combine in a single effect, and cycles in which mechanism chains form closed loops.

How to deal with the complexity vs. completeness problem?

I have placed an ordering on these hypothesis forms based on a straightforward complexity analysis of the corresponding causal structures. This ordering provides a direct means of controlling hypothesis generation: propose hypotheses of the simplest type first, those of the more complex types later. However, we want to have compelling reasons for constructing the more complex hypotheses. We do not want to generate them indiscriminately; neither do we want to exclude them entirely from consideration.

Accordingly, I have designed heuristic justification rules for each level of hypothesis. Each rule triggers on characteristic kinds of deficiencies in hypotheses at a less complex level. Hypothesis generation is explicitly controlled by permitting an incomplete hypothesis at a given level to be extended into a hypothesis at a more complex level only when the justification rule at the more complex level is satisfied.

What is the power of causal reasoning in the mechanical, electrical, and thermal domain?

Much of the work on causal reasoning about physical systems has concentrated on the digital and analog circuit domain [Barrow 84, Davis 84, de Kleer 84, Genesereth 84, Williams 84]. In this domain, the relationship be-

tween structure and behavior is well-understood because structure is nearly equated with topology and component behaviors are by design composable. In addition, circuit domain knowledge is compendious enough to make theorem proving a feasible enterprise. These properties are more elusive in the domain of mechanical, electrical, and thermal systems. The couplings between structure and behavior are less straightforward and there is no universal notation for articulating domain knowledge. Nevertheless, a number of research efforts have begun to address these problems [de Kleer and Brown 84, Forbus 84, Kuipers 84].

This thesis is in part an investigation into the power of representing and reasoning about causal relations in mechanical, electrical, and thermal systems, in the context of the difficult modelling or "black box" problem. I am interested also in isolating the contribution of causal knowledge from that of teleological knowledge. I have deliberately suppressed reasoning about the intended function of devices in the program JACK for this reason, although I fully expect that such knowledge would prove to be an additional and complementary source of constraint.

What makes for a convincing model of a device?

This is the most difficult issue which this work must address. The models of devices produced by the causal modelling system are necessarily abstractions of the real devices. The abstraction is not just an unavoidable artifact; it is an important part of the modelling process. Given the size of the hypothesis spaces being dealt with, it becomes infeasible to generate hypotheses which incorporate full details of mechanics or thermodynamics, etc., even putting aside objections concerning the completeness of any knowledge base which purports to have captured the level of detail in, say, a college physics textbook. Quite deliberately, my representations for mechanisms and constraints in the physical system domain capture abstractions of physical and causal principles.

The trick, however, is to strike a proper balance between abstraction and discriminatory power. We want suppression of detail while retaining a capability for making fine distinctions among manageably few hypotheses. We want abstractions which support sufficiently sound reasoning such that good choices can be made about which hypotheses to admit and which to prune.

An important place to look is at device models proposed by the program JACK which do not correspond to the standard designs for the given device. Ideally, these alternate models should either be genuine alternate designs for the given device, or there should be an identifiable lack of more specific knowledge which could be added to the program.

1.4 Roadmap

In Chapter 2, I offer a concise description of the causal modelling problem, I determine the sources of complexity in the problem, and I outline the approach which results in the successful modelling of several devices.

In Chapter 3, I describe the causal ontology and the set of representations I designed in support of my solution to the causal modelling problem. Chapter 4 contains a detailed description of the manifestation of that solution in the procedures which make up the program JACK.

In Chapter 5, I relate in detail the reasoning employed by the program JACK on the several device examples. For each example, I describe the generation of several plausible models and identify abstractions and shortcomings in those models.

Chapter 6 contains an analysis of the performance of the causal modelling system. Here, I offer empirical results concerning the number of hypotheses generated by the program JACK for each of the device examples. I offer also statistics concerning the effectiveness of the individual constraints and the ordering on hypotheses at constraining search. Finally, I examine the robustness of the program JACK and enumerate assumptions and limitations in my approach to the causal modelling problem.

In the final chapter, I discuss how the issues enumerated at the outset are addressed in this work, relate my research to other efforts, outline some directions for future research, and offer some ideas about potential uses for a causal modelling system.

2. The Problem: JACK in the Black Box

I refer to the problem which is the central focus of this thesis as the causal modelling problem. This term echoes the nature of the approach I have taken to solve the problem. This approach involves instantiating and composing *causal* explanations to account for the behavior of a device.

2.1 Formal Statement of the Problem

The causal modelling problem can be stated as a graph problem. The nodes of the graph correspond to the events of a device—changes in the values of its quantities. The arcs of the graph correspond to the mechanisms which map events to other events.

Some of the observable events of a device are distinguished as known inputs or primitive causes. Others are distinguished as known outputs or final effects. The task is to construct a set of directed graphs consisting of mechanisms and intermediate events which connect the known input nodes to the known output nodes. See Figure 2.1. The direction of the arcs is from cause to effect. These *causal graphs* are the output of the causal modelling system.

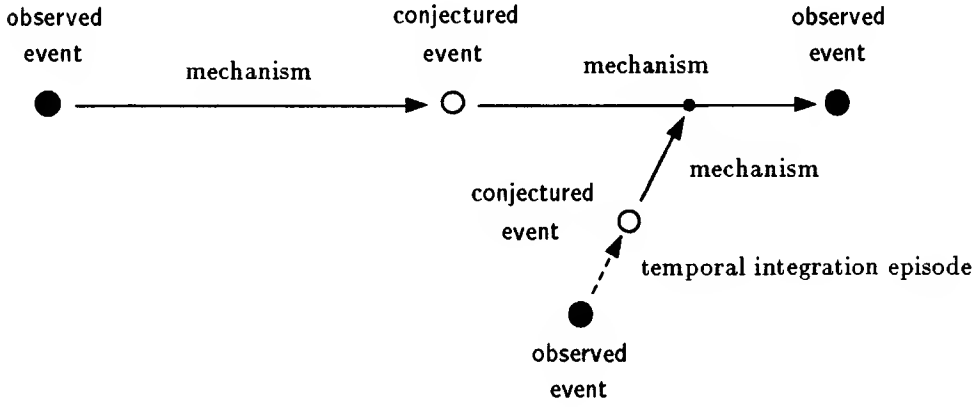


Figure 2.1. A causal graph.

The original set of observable events forms the periphery of the graphs to be constructed. The mechanisms and intermediate events correspond to hy-

potheses about what hidden mechanisms may exist and what unobservable events may take place inside the black box.

2.2 Viewpoints on the Problem

In this section, I relate causal modelling to several forms of reasoning.

2.2.1 Explanation

Causal modelling is a form of explanation. An event—a change in an observable quantity of a device—is explained in terms of other event(s) and the mechanisms which mediate between them. Motion can be explained by another motion and a mechanical coupling, by the force of gravity, by a change in temperature and thermal expansion, etc.

The explanations formed by JACK carry the *necessity* property associated with causality: If the cause(s) and mechanism(s) are present, the effect must occur. This is the basis of my use of the term *causal*; no further philosophical or other trappings should be read into it.

Forbus [Forbus 86] applies the term *measurement interpretation* to the explanation of observations of a device in terms of mechanisms and events within it. Causal modelling is measurement interpretation with the significant difference that the model of the device is not given, but is hypothesized in the very process of forming explanations.

2.2.2 Theory Formation

I take theory formation to be an elaboration of explanation; namely, explanation refined across a set of examples. The motion of the slide in the tire gauge can be explained by a coupling with a hidden object based either on attachment or simple contact. However, when the unopposed spring pushes the hidden object back into the cylinder, only the contact coupling hypothesis remains consistent with the motionlessness of the slide.

The causal modelling system verifies existing hypotheses against additional examples of behavior. This refinement of hypotheses, or theory formation, is a form of learning in the program JACK.

2.2.3 Design

Causal modelling is also an abstract form of design. The program JACK, in conjecturing compositions of mechanisms within a device to explain its

behavior, is also conjecturing how the device may have been designed.

The reasoning in the causal modelling system is qualitative, making it most like the early stages of design. There are two reasons why causal modelling falls short of full-fledged design. First, the representations for structure in the program JACK are limited. They cannot support the detailed reasoning about how structure implements function which characterizes the later stages of design. Second, the timeline of events which is one input to the causal modelling system is a specification, albeit incomplete, of the range of behavior of the device. A designer, on the other hand, does not reason directly from behavior, but works with functional specifications which are refined incrementally, and which ultimately define behavior.

2.3 The Domain

The program JACK operates in the domain of mechanical, electrical, and thermal physical systems. This class does not include electronic devices: digital, analog, or VLSI technology. The order of complexity which has been successfully tackled is roughly that of the common household gadget.

The device examples which have been implemented include a toaster, a tire gauge, an old-style bicycle drive with coaster brake, a refrigerator, and a home heating system. The program JACK models simplified versions of the more complex among these physical systems.

In the toaster example, an example of behavior in which toast turns out too light is used to recognize the role of thermal expansion in the toaster. In the pocket tire gauge example, the causal modelling system identifies the push-but-not-pull nature of the coupling which moves the slide and then leaves it there. In the bicycle drive example, the program JACK infers that independent linkages drive the rear wheel and activate the brake. The linkages operate in opposite directions so that only one can be engaged at a time. The program JACK proposes a model for a refrigerator where cooling of the interior and heating of the exterior are two halves of a cycle. Finally, a hidden fluid transport medium is offered as an explanation for the behavior of a home heating system.

2.4 Motivation

In this section, I argue for the importance of the causal modelling problem. The motivation for this investigation goes well beyond idle curiosity in modelling as an abstraction of the scientific process. There are theoretical and pragmatic reasons for studying this problem.

2.4.1 Understanding Constraint in the Physical System Domain

My primary theoretical objective in investigating the causal modelling problem is to expose sources of constraint for reasoning in the physical system domain. Such an exercise is part of a well-known and successful paradigm for conducting research in artificial intelligence [Marr 82]. Any success I achieve in identifying constraint sources can provide a starting point for other researchers working on problems in causal reasoning about physical systems.

2.4.2 Causal Models and Problem Solving

The importance of causal modelling can be argued from a purely pragmatic stance. Device models can support numerous forms of problem solving in the physical system domain, including diagnosis, monitoring, and planning. For example, in diagnosis a model of the working device can be used to recognize departures from nominal behavior, and simulation of fault models can be used to test hypotheses about the source of misbehavior.

Prediction of device behavior via model-based simulation can support decisions about how to monitor a physical system. A simulation trace generates expectations about changes in sensor values and exposes causal dependencies which can be used to assess the importance of predicted events in the continued nominal operation of the device. These assessments in turn can support decisions about how to allocate sensor resources.

Sometimes an existing model must be refined before it can support a new problem solving task. For example, a simple model of a camera might describe how the aperture width and shutter speed both contribute to exposure. However, if the goal is to take unblurred photographs, a more detailed model is needed, one which describes how aperture width affects blurring due to distance and how shutter speed affects blurring due to motion.

In Section 7.4.1, I offer scenarios which show how causal models generated by the program JACK can be used. The problem solving tasks in these scenarios are diagnosis—reasoning about faults within a device, and monitoring—the efficient allocation of sensor resources to verify the nominal operation of a device.

2.5 Analysis: How Hard is the Problem?

In this section, I conduct a straightforward worst-case complexity analysis of the causal modelling problem. The causal modelling problem is hard because the number of possible hypotheses about what mechanisms may be inside a

device is exponential in two parameters. The two sources of complexity are direct consequences of the "black box"; they arise from the lack of knowledge of the internal topology of the device. One is due to uncertainty about the lengths of the mechanism paths in the causal graph of a device; the other is due to uncertainty about how mechanism paths join in the causal graph of a device.

2.5.1 Mechanism Paths

For any pair of events, the number of possible mechanism paths between them is m^l where m is the number of possible mechanisms and l is the length of the path. See Figure 2.2. For n observable events, there are n^2 pairings of these events. The exponential contribution dominates; therefore the number of mechanism path hypotheses for a set of observable events is $\mathcal{O}(m^l)$.

Figure 2.2. Mechanism paths.

2.5.2 Mechanism Interactions

Unfortunately, linear chains of mechanisms are not the only form of hypothesis which must be considered. An effect may be the result of an interaction between multiple causes. See Figure 2.3. One cause may *enable* or *disable* another, as when the opening or closing of a valve enables or disables fluid flow. The contributions of two causes may cancel to form an *equilibrium* state, as when the forces due to air pressure and a spring balance in a tire gauge.

The number of hypotheses for interacting, or joined mechanism paths is the product of the number of hypotheses for each of the separate linear mechanism paths. The number of hypotheses for linear mechanism paths is $\mathcal{O}(m^l)$; the number of hypotheses for joined mechanism paths is therefore $\mathcal{O}(m^l m^l ... m^l)$. For p interactions, the number of mechanism hypotheses becomes $\mathcal{O}(m^{lp})$. The causal modelling problem becomes exponential in two parameters in the worst case.

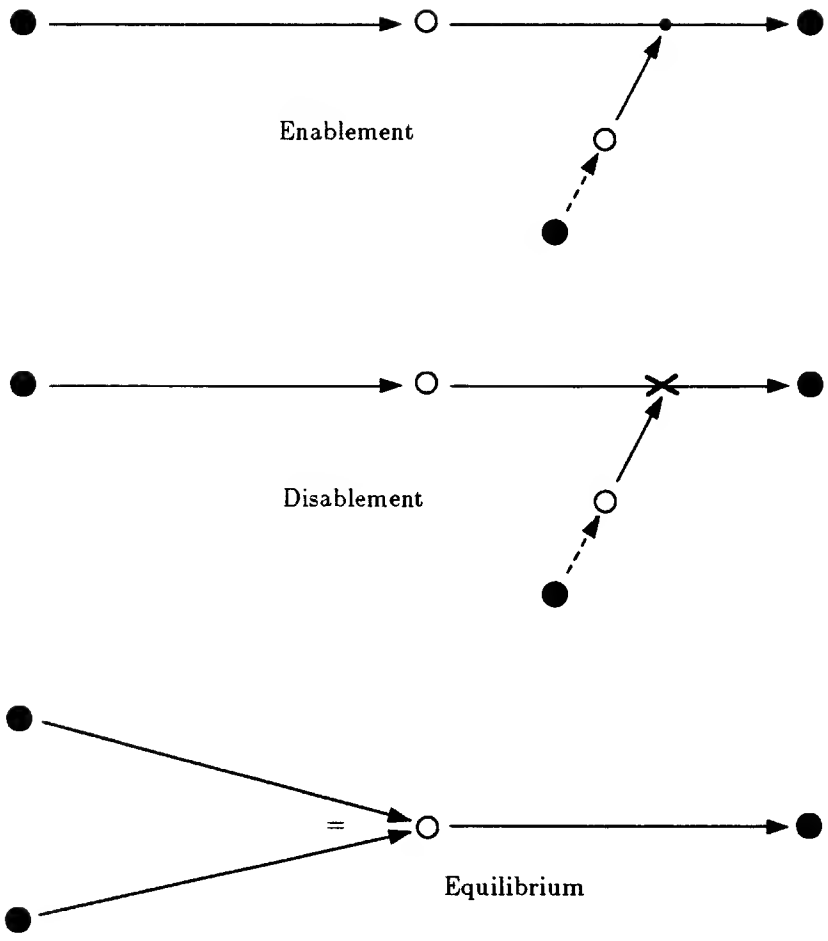


Figure 2.3. Mechanism interactions.

2.5.3 Hidden Inputs

Both linear mechanism paths and mechanism interactions may involve hidden inputs. See Figure 2.4. Although the worst-case number of ways to instantiate a given graph structure does not change in the presence of hidden inputs, the ability to constrain instantiation is compromised. All hypotheses must be consistent with the observed, incontrovertible events which make up the periphery of a causal graph. A hidden input at the periphery of a causal

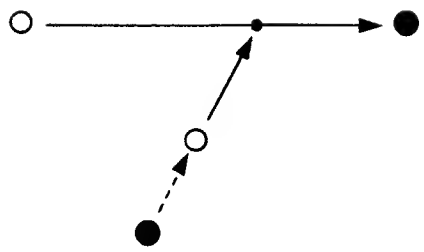


Figure 2.4. Hidden inputs.

graph is, on the other hand, a wild card which offers no source of constraint.

2.5.4 Cycles

Both linear mechanism paths and mechanism interactions may be extended to include cycles. See Figure 2.5. A cycle effectively adds an additional mechanism path to a causal graph. The worst case number of hypotheses becomes $\mathcal{O}(m^{l(p+1)})$. Although there is no new exponent, the worst case number of hypotheses for a causal graph which includes a cycle is strictly greater than the worst case number of hypotheses for the same causal graph without the cycle.

2.6 The Approach

In this section, I describe informally the constraints from the physical system domain and the ordering of hypotheses in the physical system domain which are at the crux of my solution to the causal modelling problem.

2.6.1 Physical and Causal Constraints

The constraints I enumerate here are, I believe, a fair summary of common sense concerning devices. Nevertheless, they are based in, and inherit the inviolability attributed to, physics and causality. All hypotheses about the mechanisms within devices are subject to these constraints.

The constraints concern how different observable aspects of the behavior and structure of physical systems are conserved or transformed across mecha-

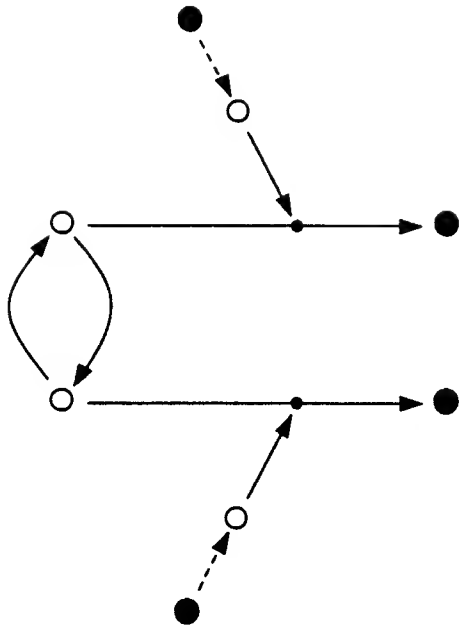


Figure 2.5. Cycles.

nisms. Any hypothesis about a configuration of mechanisms between an event taken to be a cause and an event taken to be an effect must account for any change or lack of change between the two events for all of these aspects of behavior and structure.

2.6.1.1 Type

This constraint concerns the types of quantities in a physical system. Examples of quantity types are amount and rate of position, amount and rate of temperature, amount of fluid and rate of fluid flow, etc. Proposed mechanisms must be consistent with observed type conservations or transformations between a cause and an effect.

For example, a mechanical coupling is an admissible explanation for a cause whose type is rate of position and an effect whose type also is rate of position.

2.6.1.2 Delay

The delay constraint concerns the times of occurrence of events in a physical system. Mechanism hypotheses must account for any time lag between causes and effects.

For example, electricity or a rigid coupling, whose propagation times are essentially instantaneous, are consistent hypotheses for a cause and effect which are perceptually simultaneous. Conversely, these same mechanisms cannot be offered as an explanation for events which are separated in time.

2.6.1.3 Sign

The sign constraint concerns the signs of the values of quantities in a physical system. Mechanism hypotheses must account for any conservation or transformation of sign between causes and effects.

For example, an increase in temperature can account for an increase in pressure but cannot explain a decrease in pressure. Flow in a closed system implies a decrease in amount at the cause and an increase at the effect, or vice versa. In an open system with an external source and sink, two amounts may increase or two amounts may decrease.

2.6.1.4 Direction

The direction constraint concerns the orientations in space of quantities in a physical system. Mechanism hypotheses must account for any deflection between causes and effects. The direction constraint is an elaboration of the sign constraint for vector, as opposed to scalar, quantities.

A spring, which produces a reversal in the direction of motion, is a consistent explanation for a motion followed by a motion in the opposite direction. A rigid coupling, on the other hand, which preserves orientation, is not.

2.6.1.5 Magnitude

The magnitude constraint concerns the magnitudes of the values of quantities in a physical system. Mechanism hypotheses must account for similarities or differences in magnitude between causes and effects.

For example, a rigid coupling, which transfers motion with no loss, can be a causal explanation only for motions of the same magnitude. The acceleration due to gravity over finite distances can account for velocities only within a certain range of magnitude.

2.6.1.6 Alignment

The alignment constraint concerns the relative values of quantities in a physical system. Some mechanisms require the value at the cause to be in some relation to the value at the effect. Only those mechanisms for which any such required relation is satisfied may appear in mechanism hypotheses.

For example, the direction of heat flow always is from the warmer to the cooler site. Or, stated differently, the temperature value at the cause must be greater than the temperature value at the effect. This constraint also distinguishes couplings which support pulling but not pushing, or vice versa. For example, for a non-rigid coupling such as a string, the position of the cause must be greater than the position of the effect, along the direction of motion.

2.6.1.7 Bias

The bias constraint concerns the directions of change of quantities in a physical system. Some mechanisms place a restriction on the absolute direction of change at the cause or effect. Only those mechanisms whose preferred direction of change is satisfied may appear in mechanism hypotheses.

For example, a ratchet allows motion in one direction but not in the opposite direction. A coupling based on contact, on the other hand, may engage in either direction. Condensation results from a pressure increase and evaporation results from a pressure decrease.

2.6.1.8 Displacement

The displacement constraint concerns the locations of objects in a physical system. Mechanism hypotheses must account for any physical separation between causes and effects.

For example, thermal expansion cannot account for a temperature change in one physical object and a motion in another because thermal expansion takes place entirely within one physical object. However, thermal expansion preceded by a heat flow, or thermal expansion followed by a mechanical coupling can explain the observation because in both cases, the additional mechanism is sufficient to account for the change in location.

2.6.1.9 Medium

The medium constraint concerns the connections between objects in a

physical system. Examples of connections are attachment in a rigid coupling, an unobstructed line-of-sight path in radiative heat flow, etc. Mechanisms whose associated connection type cannot be established or conjectured between the sites of causes and effects may not appear in mechanism hypotheses.

For example, gas flow is an admissible hypothesis when two physical objects are joined, but is untenable when they are separated. A valve must span a conduit in order to explain a change in flow.

2.6.2 An Ordering on Hypotheses

Figure 2.6 shows an ordering on types of device model hypotheses. These hypothesis types correspond to different causal graph structures. The ordering is a partial ordering. Linear mechanism paths may be extended into branching mechanism interactions. Either of these hypothesis types may involve hidden inputs. Finally, hidden input hypotheses may be extended to include cycles.

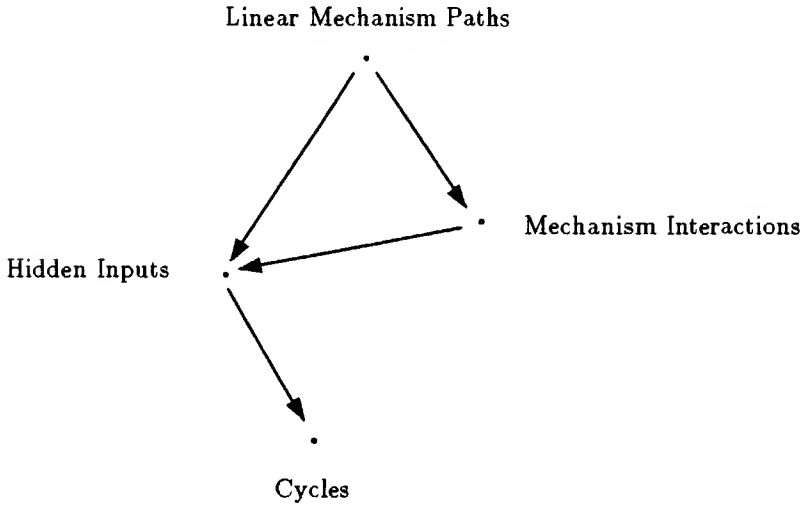


Figure 2.6. The ordering on hypotheses.

The directions of the arcs in this ordering are based on either: (1) in the case of arcs leading to the hidden input node, loss of constraint due to unobservable events on the periphery of the causal graph, or (2) in the case of all other arcs, a hypothesis space of strictly greater cardinality.

The causal modelling system must not proceed through this ordering indiscriminately. There should be clear justifications for moving from one level of hypothesis construction to another, for each level jump implies an explosion. I have designed a set of heuristic look-ahead rules for justifying level jumps. These rules capture manifestations of the following principle: *Incomplete hypotheses often exhibit characteristic deficiencies.* These signatures can indicate into what other form of hypothesis a deficient hypothesis should be extended. Incomplete hypotheses displaying these signatures have a heuristically justified chance of being reparable at another level.

These heuristics quite deliberately prevent the program JACK from making a complete search. In fact, some of the arcs in the ordering on hypotheses are suppressed entirely. Only certain deficient hypotheses ever are extended into hypotheses at another level; no successful hypothesis ever is extended. This focusing is necessary to offset the explosion awaiting at more complex or less constrained levels of hypothesis generation.

2.6.2.1 Mechanism Paths

Linear mechanism paths form the root of the hypothesis ordering. The program JACK always initiates hypothesis construction at this level; no justification is needed.

2.6.2.2 Mechanism Interactions

Two kinds of mechanism interaction can be distinguished: (1) enablements and disablements where one mechanism renders another active or inactive, and (2) equilibria where the contributions of separate mechanisms cancel each other.

An unsuspected enablement situation can be characterized by an unexplained delay; the expected effect occurs, but too late. In addition, the magnitude of the effect may be less than expected. An example is the opening of a valve and an increasing amount of fluid after a fluid source has been already established.

The hallmark of a disablement situation is an arrested change occurring after the expected appearance of a non-zero effect. An example is the closing of a valve and a resulting stable amount of fluid after a fluid source has become already manifest in an increasing amount of fluid.

The signature of an equilibrium situation is the same as that for a disablement situation: an unexpected return to zero after an expected effect has

occurred. An example is the establishment of a fluid sink and a resulting stable fluid level after a fluid source has become already manifest in an increasing amount of fluid.

2.6.2.3 Hidden Inputs

Certain mechanisms can appear only in enablement or disablement interactions; they cannot stand alone along linear mechanism paths. The valve which permits or inhibits fluid flow is an example. The opening or closing of the valve does not directly cause the genesis or cessation of a fluid flow. Nothing happens in the absence of a fluid source.

A hidden input situation is signalled by any linear mechanism path hypothesis which includes one of the non-stand-alone mechanisms. These hypotheses always incorrectly predict zero effects because of the missing fluid, current, heat, etc. sources.

2.6.2.4 Cycles

One possible signature of cycles within a device is repetitive patterns of behavior. However, cyclic behavior may not manifest at the macroscopic level; the iterations may be blurred into apparently continuous changes. For example, the cooling of the interior of a refrigerator, perceived to be continuous, is actually the result of repeated cooling pulses supplied by the evaporation half of the refrigerant cycle.

A more perspicuous signature for cycles can be derived by noting that hidden inputs imply unknown sources and sinks. Well-designed devices are expected to have a minimum of sources and sinks because conservation laws demand that sources must be explicitly supplied and sinks must be explicitly removed. Potential sources and sinks within a device can be avoided by forming cycles where gains in one part of the cycle are offset with losses in another part. A signature for this kind of synergistic cycle is the presence of at least one conjectured source and one conjectured sink.

3. Representations and Ontology: The World According to JACK

Supporting my approach to the causal modelling problem is a host of representations and an ontology—partially inherited and modified from the work of other researchers, partially designed by myself. The representations are aimed at supporting reasoning about the behavior and structure of devices and at exposing the constraints which prevent mechanisms from being arbitrarily composed. The ontology is aimed at enumerating the types of causal graph structures for devices, and the primitives which make up causal graphs.

3.1 Quantities

Quantities are continuous properties of physical objects. Examples of quantities from the devices modelled by the program JACK include the position of the lever outside a toaster, the amount of gas inside a tire, the angular velocity of the pedal in a bicycle drive, and the temperature inside a refrigerator, or a hot-water radiator.

My representation for quantities is similar to and was inspired by the quantity representation designed by Forbus in his Qualitative Process Theory [Forbus 84]. Quantities are represented as triples of the form:

{·QUANTITY· *physical-object property order*}

The quantities enumerated in the last paragraph are represented as:

{·QUANTITY· *Lever Position Rate*}
{·QUANTITY· *Tire Amount-of-Gas Amount*}
{·QUANTITY· *Pedal Angle Rate*}
{·QUANTITY· *Interior Temperature Amount*}
{·QUANTITY· *Radiator Temperature Amount*}

The first slot in the representation of a quantity denotes the physical object whose continuous property is being described. The second slot denotes the type of continuous property being described. The third slot denotes whether the property itself or its first derivative is being described.

3.1.1 Types

The type of a quantity is defined as its property and order. Thus, for example, displacement {·TYPE· *Position Amount*} and velocity {·TYPE· *Position Rate*} are treated as distinct types.

3.1.2 Derivatives

The order of a quantity is either and only *Amount* or *Rate*; higher-order derivatives cannot be represented explicitly.

3.1.3 Values

The values of quantities are represented in two complementary fashions: as discrete qualitative regions and as ranges of orders of magnitude. In many cases, the distinctions afforded by qualitative regions are sufficient to draw useful inferences—that the value of a quantity is or is not changing, that a switch is open or closed. Ranges of orders of magnitude allow finer distinctions to be made, for example between the magnitude of motion due to gravity and the magnitude of motion due to the uncoiling of a spring. Ranges of orders of magnitude represent a compromise between purely symbolic qualitative reasoning and precise quantitative reasoning.

3.1.3.1 Qualitative Regions

Qualitative regions are represented by symbols such as *Positive*, *Down*, and *Ambient*.

3.1.3.2 Order of Magnitude Ranges

Ranges of order of magnitude are represented as:

$\{\cdot\textsc{range}\cdot\ radix^{min} : radix^{max}\}$

The low end of an order of magnitude range is the base *radix* raised to the exponent *min* and the high end is the base *radix* raised to the exponent *max*. For example, the numeric interval $[0.1 : 100.0]$ could be represented as $\{\cdot\textsc{range}\cdot\ 10^{-1} : 10^{2}\}$ or as $\{\cdot\textsc{range}\cdot\ 2^{-3} : 2^{7}\}$.

The qualitative value *Positive* is associated with the order of magnitude range $\{\cdot\textsc{range}\cdot\ 2^{-\infty} : 2^{\infty}\}$; where $2^{-\infty}$ is a limiting value corresponding to zero and 2^{∞} is a limiting value corresponding to infinity. Within this range, many distinctions can be made which cannot be made in the associated qualitative region. For example, the rate of motion due to gravity might be in the range $\{\cdot\textsc{range}\cdot\ 2^{0} : 2^{8}\}$; while motion due to a spring might be in the range $\{\cdot\textsc{range}\cdot\ 2^{0} : 2^{12}\}$. An observed motion in the range $\{\cdot\textsc{range}\cdot\ 2^{10} : 2^{10}\}$ would exclude the gravity hypothesis. The qualitative region *Positive* hides the distinction which supports this inference.

3.1.4 Value Spaces

Quantities take on values from a continuous range. This range is termed the value space of the quantity and is represented as a list of qualitative regions. For example, the value space of the refrigerator quantity {·QUANTITY· *Interior Temperature Amount*} is {·VALUE SPACE· *Cold Ambient*}; the value space of its derivative {·QUANTITY· *Interior Temperature Rate*} is {·VALUE SPACE· *Negative Zero Positive*}.

Value spaces are total orderings on the values of a quantity. Inequality relations between values of a single quantity can be inferred directly from the positions of the values in the value space list. Forbus describes how the value space (quantity space) of a single quantity may sometimes be a partial order. However, none of the device examples on which the program JACK was tested included quantities with partially ordered value spaces.

3.1.5 Zeros

One value in the value space of a quantity is distinguished as the zero value for that quantity. The sign of any value of a quantity can be inferred from the position of that value in the value space relative to the zero value. When no zero value is declared for a quantity, all values are assumed to be positive. For example, no zero value is declared for the quantity {·QUANTITY· *Interior Temperature Amount*}; both values in its value space {·VALUE SPACE· *Cold Ambient*} are positive.

3.1.6 Inequalities

Explicit *Less*, *Greater*, or *Equal* relations may be asserted between values in the value spaces of different quantities:

{·RELATION· *Warm Greater Cold*}

Greater relations are asserted automatically between contiguous values in the value spaces of single quantities. These relations also contribute to partial orders among the values of quantities. These partial orders enable inferences to be drawn about inequality relations between the values of different quantities. See Figure 3.1.

3.1.7 Orientations

The orientation of a quantity is the direction of increasing values in its value space. Quantities may be defined in different coordinate systems. In this

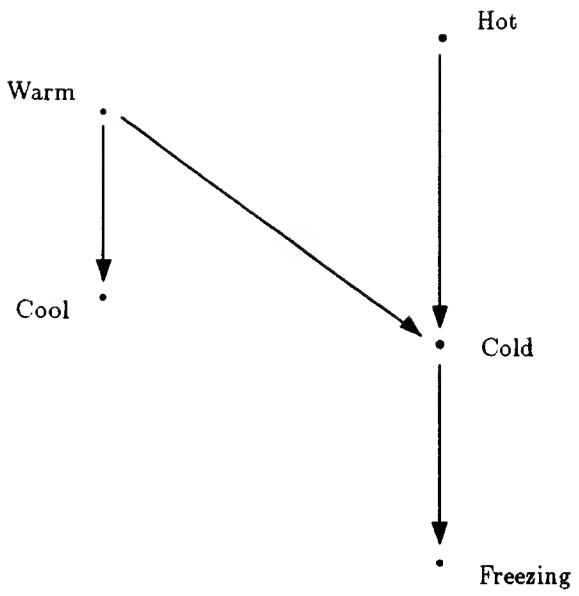


Figure 3.1. A partial order on quantity values.

case, the relative orientation of the two systems is needed to project a change in the value of one quantity onto the value space of another. See Figure 3.2. Even scalar quantities may have opposite orientations; in this case an increase in one is in the same direction as a decrease in the other, and vice versa.

Quantities are assigned orientations through an explicit relation:

{·RELATION· {·QUANTITY· *Lever Position Amount*} *Orientation* $+Y$}

Relative orientations also are defined through relations:

{·RELATION· $+X$ *Opposite* $-X$}
{·RELATION· $+X$ *Perpendicular* $+Y$}
{·RELATION· $+X$ *Skewed Cylinder-Axis*}

3.2 Relations

Structural relations among physical objects, geometrical relations among orientations, and inequality relations among the values of quantities all are represented with the form:

{·RELATION· *subject relation object*}

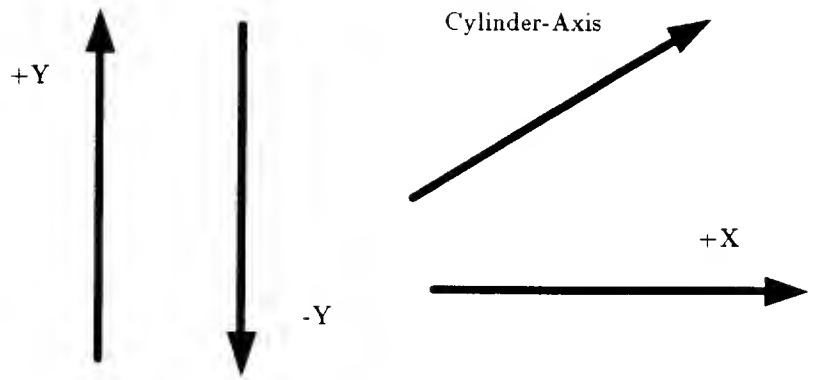


Figure 3.2. Orientations.

Examples of relations are:

{·RELATION· *Warm Greater Cold*}
{·RELATION· *+X Opposite -X*}
{·RELATION· *Lever Attached-To Carriage*}

Arbitrary structures may appear in the subject and object slots of relations, including quantities and other, nested relations:

{·RELATION· *-Y Orientation-Of* {·QUANTITY· *Earth Gravity Amount*}}

The value space for all relations is {·VALUE SPACE· *True Unknown False*}. Unlike the value spaces for quantities, there is no ordering placed on the possible values for relations, nor are there order of magnitude intervals associated with these truth values.

3.2.1 Inverse Relations

Many types of relation have inverses. For example, the inverse of *Greater* is *Less*, the inverse of *Attached-To* is *Attached-To*. Inverse relations are automatically asserted whenever relations with defined inverses are asserted. For example, the assertions:

{·RELATION· *Warm Greater Cold*}
{·RELATION· *Lever Attached-To Carriage*}

result in the additional assertions:

{·RELATION· *Cold Less Warm*}

{·RELATION· *Carriage Attached-To Lever*}

3.3 Time

The observations of physical systems which the causal modelling system attempts to explain describe how the values of quantities and relations change over time. Representations and methods for temporal reasoning have been a major focus of research recently [Allen 83, Vere 83, Shoham 86, Williams 86, Dean and McDermott 87]. I have adopted a representation for time which inherits results from these efforts.

3.3.1 Intervals

My temporal representation is based on intervals. An interval has a beginning and an end.

{·INTERVAL· *beginning* : *end*}

The beginning and end of an interval are primitive intervals called moments. The beginning and end of a moment is always the moment itself. Thus the interval {·INTERVAL· 10 : 10} and the moment {·MOMENT· 10} are the same interval. Contiguous intervals, such as {·INTERVAL· 3 : 4} and {·INTERVAL· 4 : 9}, meet at instants, or temporal points. See Figure 3.3.

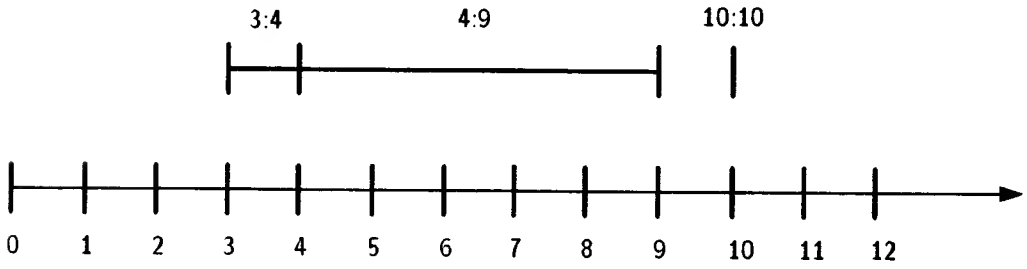


Figure 3.3. Intervals and Moments.

3.3.2 Histories

Interval-value pairs add the temporal dimension to assertions about the values of quantities and relations. For example, an assertion that the value of the

quantity {·QUANTITY· *Interior Temperature Amount*} is *Ambient* during the interval {·INTERVAL· 0 : 60} is represented by associating with this quantity the interval-value pair:

[0 : 60 *Ambient*]

A *history* is a list of interval-value pairs which describe changes in and durations of the values of quantities and relations. All quantities and relations have histories. For example, the history of the relation {·RELATION· *Cylinder Joined-To Tire*} might be:

{·HISTORY· [0 : 59 *False*] [60 : 69 *True*][70 : $+\infty$ *False*]}

The values of quantities and relations always are assumed to be *persistent.* They are changed only by explicit subsequent events.

3.4 Behavior

The set of constraints which contribute to my solution to the causal modelling problem may be divided among three classes: those which have to do with, respectively, the types of quantities, the behavior of devices, and the structure of devices. Quantity types are described in Section 3.1.1. The constraints which treat device behavior are delay, sign, direction, magnitude, alignment, and bias.

3.4.1 Delay

Time lags between cause and effect are represented by ranges of order of magnitude. For example, a delay of one second to one minute, i.e., the interval {·INTERVAL· 1 : 60}, is represented as {·RANGE· 2^0 : 2^6}. Temporal intervals derived from the observation of a device, where time is treated on a linear scale, are converted to order of magnitude ranges before being used in reasoning about delays.

3.4.2 Sign

The signs of quantities are represented by qualitative regions. The set of possible signs is any subset of {*Negative Zero Positive*}. For example, a non-zero sign is represented as {*Negative Positive*}.

3.4.3 Direction

Deflections in orientation between cause and effect also are represented by qualitative regions. The set of possible directions is any subset of {*Parallel*

Opposite Perpendicular Skewed}. See Figure 3.4. The most ambiguous deflection between scalar quantities is {*Parallel Opposite*}.

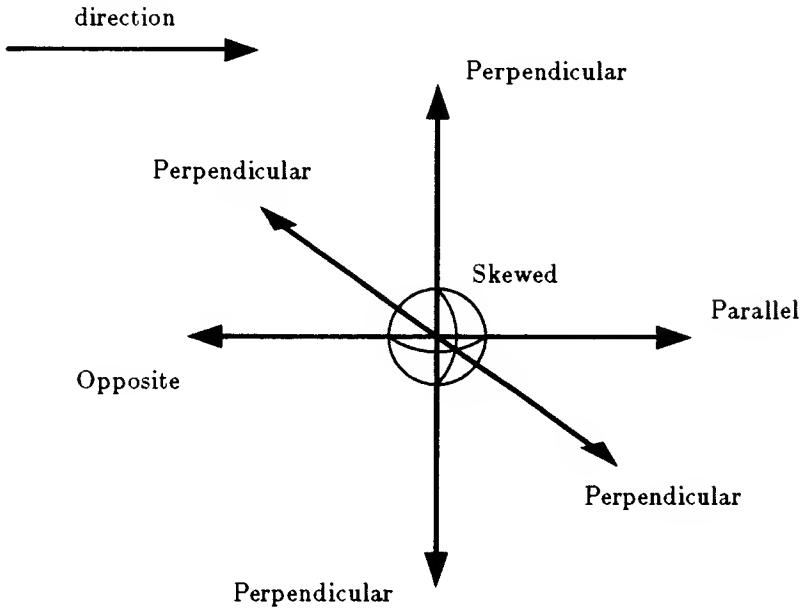


Figure 3.4 Qualitative Directions.

3.4.4 Magnitude

The magnitudes of the values of quantities are represented by order of magnitude ranges. For example, a magnitude between -10 and $+100$ is represented as {·RANGE· $2^{-\infty} : 2^{7}$}, where $2^{-\infty}$ corresponds to zero. Note that information about sign is suppressed here. Sign and magnitude are treated separately.

3.4.5 Alignment

Constraints on the relative values at cause and effect are represented directly by inequality relations. The set of possible inequality relations is any subset of {*Less Equal Greater*}. For example, an alignment of {*Greater*} requires that the value at the cause be greater than the value at the effect, as in {·RELATION· *Warm Greater Cold*}.

3.4.6 Bias

Required directions of change at cause and effect are represented by bias relations. The set of possible bias relations is any subset of {*Down-Down Down-Up Up-Down Up-Up*}. For example, a bias of {*Up-Down Down-Down*} constrains the effect to decrease without restricting the direction of change at the cause. A bias of {*Down-Down Up-Up*} is equivalent to a direct dependence between the quantity at a cause and the quantity at an effect. See Figure 3.5.

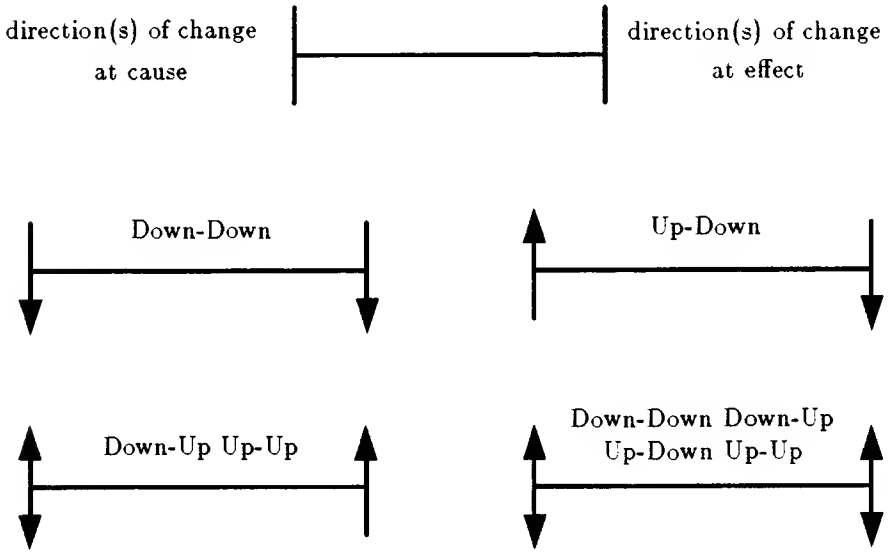


Figure 3.5. Bias relations.

3.5 Structure

The constraints which treat device structure are displacement and medium.

3.5.1 Displacement

Physical distances between cause and effect are represented by qualitative regions. The set of possible displacements is any subset of {*Same Different*}.

The displacement of a physical object with itself is *Same*; with any other physical object is *Different*.

3.5.2 Medium

Physical connections between causes and effects are represented by structural relations. Examples of structural relation types are *Attached-To*, *Touches*, *Line-of-Sight-To*, and the identity structural relation *Same*. Structural relations may change over time so that a relation such as {·RELATION· *Tire Joined-To Cylinder*} may be *True* only intermittently.

3.6 Observations

Observations of devices consist of assertions about changes in quantities and relations over time and along with several declarations. The declarations describe: (1) observable physical objects of the device, (2) observable quantities of those physical objects, (3) the value spaces of those quantities with individual values stated both as symbolic regions and as order of magnitude ranges, (4) the zero values of those quantities, (5) inequalities between the values of different quantities, and (6) scales associated with quantities and relative orientations among those scales.

The observations of a toaster, tire gauge, bicycle drive, refrigerator, and home heating system input to the program JACK appear in Appendix A.

3.6.1 Events

Events are changes in the values of quantities and relations. Events are represented by the quantity or relation involved, the new value achieved, and the moment at which the change took place. Examples of events are:

{·EVENT· *Lever Position Amount Down 61*}
{·EVENT· *Tire Joined-To Cylinder False 70*}

Events are recorded in the histories of quantities and relations. For example, the events:

{·EVENT· *Lever Position Amount Up 0*}
{·EVENT· *Lever Position Amount Down 61*}
{·EVENT· *Lever Position Amount Up 187*}

are recorded in the history of the quantity {·QUANTITY· *Lever Position Amount*} as:

{·HISTORY· [0 : 60 *Up*] [61 : 186 *Down*][187 : $+\infty$ *Up*]}

3.6.2 Timelines

The events of an observation are collected into a timeline. See Figure 3.6. Timelines are moment-centered indexings of events; they complement histories, which are quantity-centered or relation-centered indexings of events.

	0:00	1:00	1:00.1	1:00.2	2:00	2:00.1
Slide Position Amount	G0			G28		G0
Slide Position Rate	Zero		Positive	Zero	Negative	Zero
Tire Amount-of-Gas Amount	P28			P28		
Tire Amount-of-Gas Rate	Zero	Negative		Zero		
Earth Gravity Amount	G					
Earth Gravity Rate	Zero					

Figure 3.6. A timeline.

3.7 Mechanisms

Mechanisms are the building blocks for forming causal explanations. Mechanisms are represented exactly by how they impose restrictions on type, behavior, and structure. Each mechanism is defined in terms of the constraints for type, delay, sign, direction, magnitude, alignment, bias, displacement, and medium.

3.7.1 Constraints on Type, Behavior, and Structure

Every mechanism has a specific quantity type associated with its cause and with its effect. For example, the cause type for the mechanism *Condensation* is {·TYPE· *Pressure Rate*}; its effect type is {·TYPE· *Temperature Rate*}. This mechanism can explain only this particular type transformation.

The time constant of a mechanism determines the range of delays it can account for. For example, the time constant associated with mechanisms such

as *Electricity*, *Rigid-Coupling*, and *Gravity* is $\{\cdot\text{RANGE}\cdot\ 2^{\infty} : 2^{\infty}\}$. Delays are computed by dividing the time constant of a mechanism into the characteristic distance for a device. These mechanisms cannot explain any non-zero delay.

The sign of the quantity dependence associated with a mechanism restricts the sign conservations or transformations it can explain. For example, the sign associated with the *Conductive-Heat-Exchange* mechanism is *Negative*. This mechanism can explain only causes and effects of opposite sign.

The deflection associated with a mechanism determines the changes of direction it can account for. For example, the deflection associated with the *Spring* mechanism is *Opposite*. This mechanism can account for only reversals of direction.

The efficiency of a mechanism determines what changes in magnitude it can explain. For example, the *Electro-Thermal* mechanism, which subsumes a range of electrical resistances, has an efficiency of $\{\cdot\text{RANGE}\cdot\ 2^{-7} : 2^{7}\}$ and can explain a range of temperature changes. On the other hand, the *Rigid-Coupling* mechanism has perfect efficiency ($\{\cdot\text{RANGE}\cdot\ 2^{0} : 2^{0}\}$) and cannot explain any change in magnitude. Efficiency as defined here includes the notion of advantage; efficiencies may be greater than one.

The alignment relation associated with a mechanism places a restriction on the relative values at cause and effect. For example, the *Non-Rigid-Coupling* mechanism has an alignment of *Greater* and is inconsistent with an observation where the position of the cause is less than the position of the effect, along the direction of motion.

The bias relation of a mechanism constrains the directions of change at cause and effect. For example, the *Electro-Thermal* mechanism, which has a bias of {*Down-Up Up-Up*} can explain only increases in temperature.

The distance associated with a mechanism determines the displacements between cause and effect it can account for. For example, mechanisms such as *Thermal-Expansion* and *Electro-Mechanical* have an associated distance of *Same* and cannot explain interactions between events at different physical objects.

The medium associated with a mechanism indicates the structural relation which must obtain between cause and effect. For example, the medium associated with the *Contact-Coupling* mechanism is *Touches*. This mechanism can appear in a causal explanation for two motions unless a *Touches* relation between the physical object associated with the cause and the physical object associated with the effect is known to be *False*.

3.7.2 Vocabulary of Mechanisms

The complete definitions for the mechanisms *Rigid-Coupling* and *Thermal-Expansion* are:

(DefMechanism *Rigid-Coupling*
:cause-type {·TYPE· *Position Rate*}
:effect-type {·TYPE· *Position Rate*}
:distance *Different*
:time-constant {·RANGE· $2^{\infty} : 2^{\infty}$}
:sign *Positive*
:deflection *Parallel*
:efficiency {·RANGE· $2^{0} : 2^{0}$}
:alignment {*Less Equal Greater*}
:bias {*Up-Up Down-Down*}
:medium *Attached-To*)

(DefMechanism *Thermal-Expansion*
:cause-type {·TYPE· *Temperature Rate*}
:effect-type {·TYPE· *Position Rate*}
:distance *Same*
:time-constant {·RANGE· $2^{\infty} : 2^{\infty}$}
:sign *Positive*
:deflection {*Parallel Opposite Perpendicular Skewed*}
:efficiency {·RANGE· $2^{-17} : 2^{-10}$}
:alignment {*Less Equal Greater*}
:bias {*Up-Up Down-Down*}
:medium *Same*)

The entire vocabulary of mechanisms provided to the program JACK appears in Appendix B.

3.8 Causal Graphs

The program JACK constructs causal graphs which relate observable events of a device. These graphs represent hypotheses about hidden configurations of mechanisms whereby events cause other events. The nodes of these graphs correspond to device events; the arcs correspond to mechanisms.

3.8.1 Linear Mechanism Paths

The simplest kind of causal graph is a linear chain of mechanisms. An example is in Figure 3.8. Mechanism paths are represented by (1) the list of

mechanisms which make up the path, (2) the cause event which initiates the path, (3) the effect event which terminates the path, and (4) a list of event nodes which describe how type, delay, sign, direction, magnitude, alignment, bias, displacement, and medium are constrained by the mechanisms along the path. The representation of the mechanism path in Figure 3.7 (suppressing the event nodes for brevity) is:

{·MECHANISM PATH·
 :mechanisms (*Electricity Electro-Thermal*)
 :cause-event {·EVENT· *Outlet Charge Rate Positive 0*}
 :effect-event {·EVENT· *Coils Temperature Rate Positive 61*}}

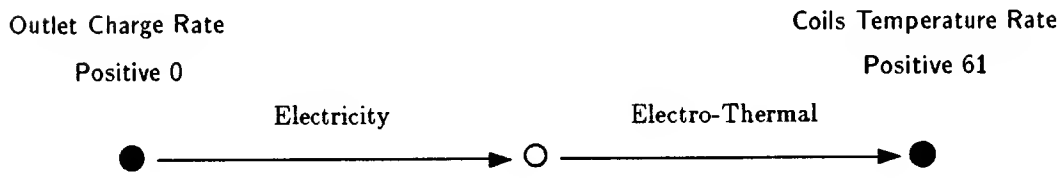


Figure 3.7. A mechanism path.

3.8.2 Event Nodes

Device events also are represented in terms of the constraints for type, delay, sign, direction, magnitude, alignment, bias, displacement, and medium. The values propagated along a mechanism path for each of these constraints make up a detailed description of the events which are expected to take place along the path. For example, the node describing the expected event between the *Electricity* and *Electro-Thermal* mechanisms in the mechanism path of Figure 3.8 is:

{·EVENT NODE·
:type {·TYPE· *Charge Rate*}
:delay {·RANGE· $2^{-\infty} : 2^{-\infty}$}
:sign {*Positive*}
:direction {*Parallel Opposite Perpendicular Skewed*}
:magnitude {·RANGE· $2^{-7} : 2^{3}$}
:alignment {*Less Equal Greater*}
:bias {*Positive*}
:displacement {*Different*}
:medium {*Coils*}}

The first event node in a mechanism path is derived from the initial cause; the last event node must be compatible with the final effect.

3.8.3 Mechanism Interactions

Linear mechanism paths are merely the simplest kind of causal graph. Mechanism paths also may join to form branching graph structures. See Figure 3.8. Mechanism interactions are represented by mechanism paths in which other mechanism paths are embedded. For example, the interaction of Figure 3.8 is represented as:

{·MECHANISM PATH·
:mechanisms (
Electricity
{·MECHANISM PATH· *Integration Switch*}
Electro-Thermal)
:cause-event {·EVENT· *Outlet Charge Rate Positive 0*}
:effect-event {·EVENT· *Coils Temperature Rate Positive 61*}}

Representations of mechanism interactions reflect the order in which hypotheses are generated. In this example, the linear mechanism path {·MECHANISM PATH· *Electricity Electro-Thermal*} was considered before the interaction.

3.8.4 Hidden Inputs

Mechanism paths may involve hidden inputs. See Figure 3.9. Hidden input paths are represented as are other mechanism paths, except that a cause event is missing. The representation of the hidden input path in Figure 3.9 is:

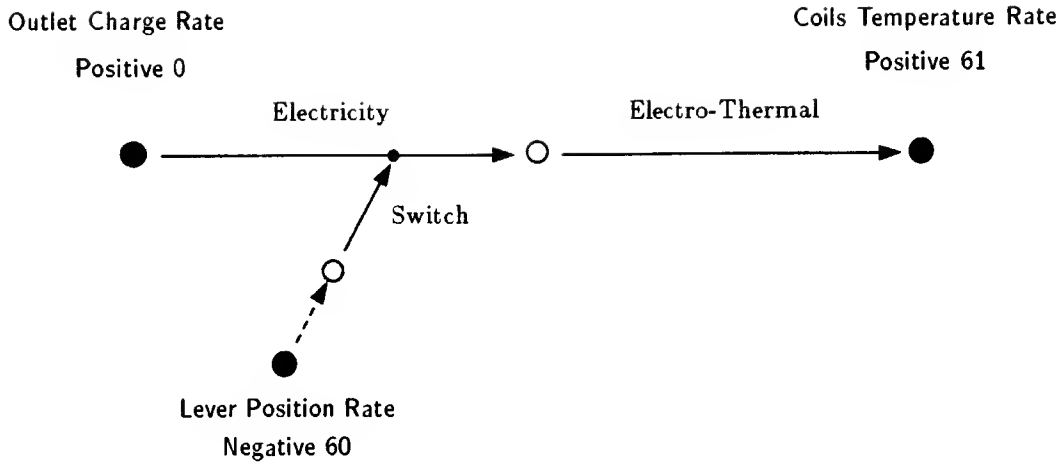


Figure 3.8. A mechanism interaction.

{·MECHANISM PATH·
:mechanisms (
Energy-Exchange
{·MECHANISM PATH· *Electricity Electro-Mechanical Compression Integration Condensation*})
:cause-event nil
:effect-event {·EVENT· *Exterior Temperature Rate Positive 61*}}

3.8.5 Cycles

Causal graphs also may involve cycles. See Figure 3.10. A cycle is represented by the list of hidden input paths which form the two halves of the cycle. There is no initial cause event or final effect event. The cycle of Figure 3.10 is represented as:

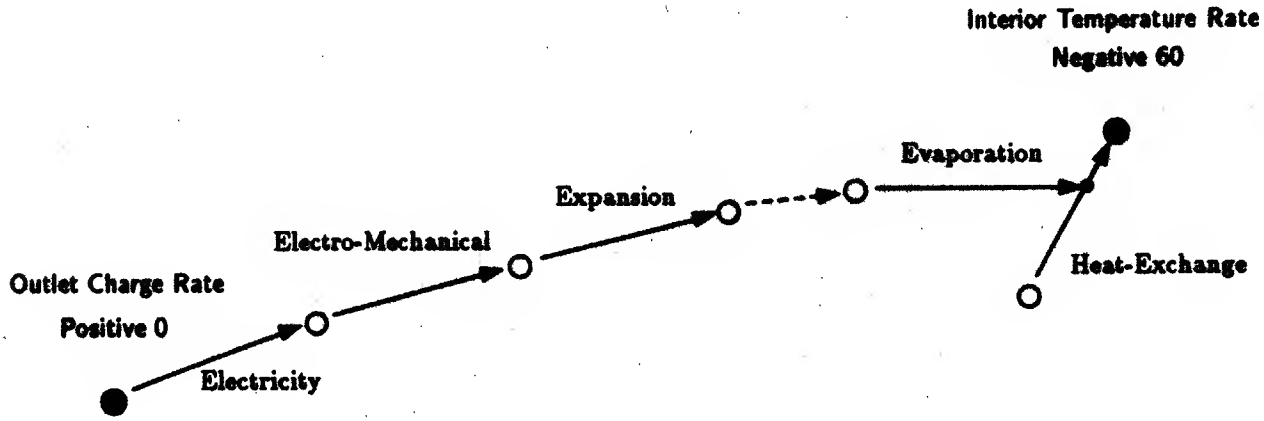


Figure 3.9. A hidden input path.

```
{·MECHANISM PATH·
 :mechanisms (
     (Energy-Exchange
      {·MECHANISM PATH· Electricity Electro-Mechanical
          Compression Integration Condensation})
     (Energy-Exchange
      {·MECHANISM PATH· Electricity Electro-Mechanical
          Expansion Integration Evaporation}))
 :cause-event nil
 :effect-event nil}
```

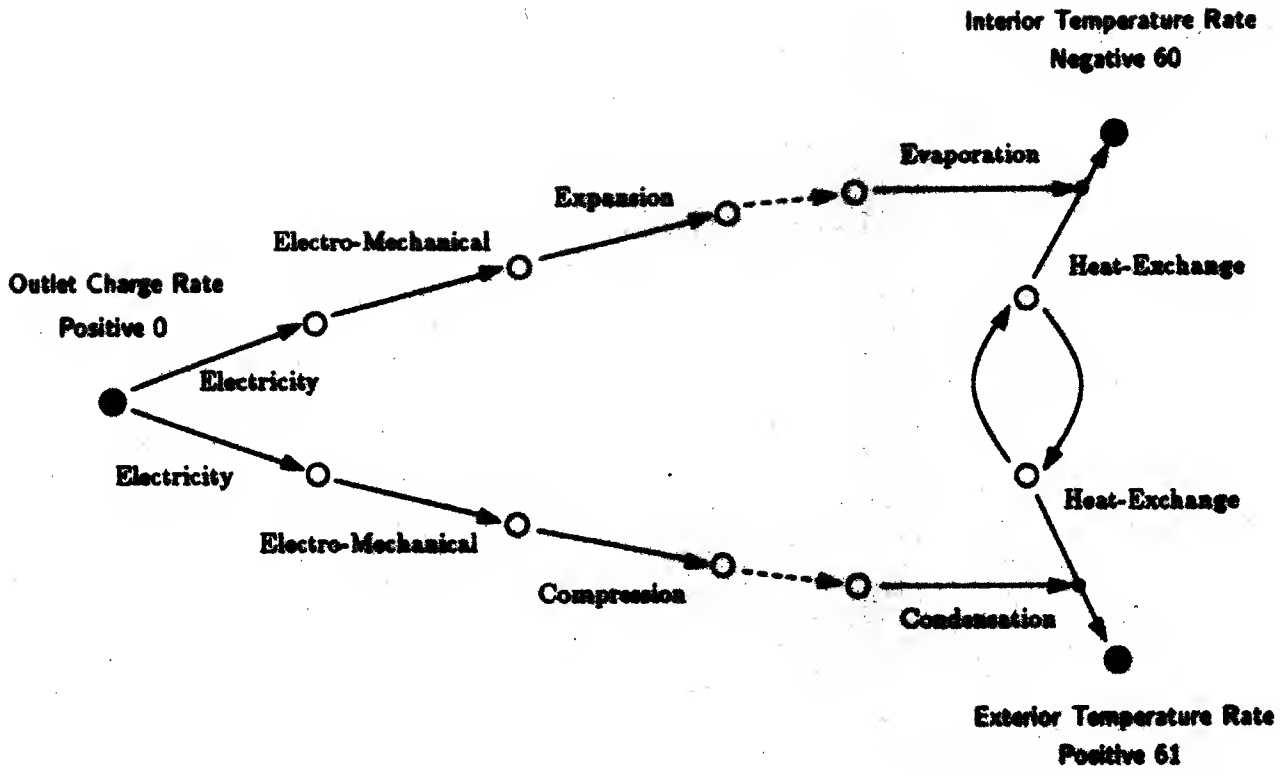


Figure 3.10. A cycle.

4. The Procedures: JACK of Some Trades

In this chapter, I describe the procedures which generate manageably sized sets of hypotheses about hidden mechanisms within devices. These procedures have been implemented in the program called JACK. JACK is, of course, an acronym and like most acronyms its interpretation was determined after its conception. JACK stands for Justified Assertion of Causal Knowledge. JACK is coded in Zetalisp and runs in Genera 7.1 for Symbolics series 3600 LispMachines.

4.1 The Causal Modelling Procedure

There are two inputs to the causal modelling procedure at top level: an observation of a device and a vocabulary of mechanisms. The output of the causal modelling procedure is a set of causal graphs which represent hypotheses about hidden configurations of mechanisms which can explain the observed events. The causal modelling procedure at top level is shown in Procedure 4.1. Details are added to this procedure throughout this chapter.

Given: a timeline, a mechanism vocabulary,
  a maximum mechanism path length l_{max},
  and a maximum number of interactions p_{max}
For each $(p+1)$-tuple of events from the timeline
  consisting of up to p_{max} causes and one effect,
  with no cause taking place after the effect
  Generate hypotheses of up to p_{max} interacting mechanism
    paths of up to length l_{max}

Procedure 4.1. Top-level causal modelling procedure.

4.1.1 Causal and Qualitative Simulation

Each generated hypothesis forms part of a proposed device model. Each partial model is simulated by propagating and combining values for the constraints on type, behavior, and structure along the proposed mechanism paths. Predicted values describe expected events which must be compatible with observations for a hypothesis to be admitted. See Procedure 4.2.

Given: a set of cause events, an effect event,
  and a hypothesized causal graph which connects them.
For each constraint
  Derive seed value(s) from cause event(s)
  Derive observation from effect event
  Propagate and combine values along the mechanism paths
    in the causal graph
  Compare propagation to observation
  When propagation and observation are compatible
    Admit hypothesis

Procedure 4.2. Testing of hypotheses via simulation.

The propagation of each constraint along a proposed mechanism path is seeded by values derived from the event taken to be the cause. The type constraint is seeded with the type of the quantity at the cause event. The seed value for the delay constraint is the zero value $\{\cdot\textbf{RANGE}\cdot\ 2^{-\infty} : 2^{-\infty}\}$. The propagation of signs is initialized with the sign of the cause event. The direction constraint is seeded with the value *Parallel.* The initial value for the magnitude constraint is the magnitude of the value associated with the cause event. The initial alignment is always $\{Less\ Equal\ Greater\}$. The bias constraint is seeded also with the sign of the cause event. The propagation of displacement proceeds from the value *Same.* Finally, the medium constraint is initialized with the physical object associated with the cause event.

The values propagated along a proposed mechanism path for each constraint are verified against the actually observed device event taken to be the effect. The pertinent observation for the type constraint is the type of the quantity at the effect event. The observed delay is the difference in time between the cause event and the effect event. The target sign is the sign of the effect event. The observed change of direction between cause and effect is the relative orientation between the quantity of the cause event and the quantity of the effect event, propagated with the value *Opposite* if the events are of opposite sign. The target magnitude is the magnitude of the effect event. The observed alignment is computed from the partial order, if any, relating the values at cause and effect. The target bias also is the sign of the effect event. The observed displacement between cause and effect is *Same* if the physical objects associated with the cause event and the effect event are the same; otherwise *Different.* Finally, the target value for the medium constraint is the physical object associated with the effect event.

4.1.2 Propagation Rules

Values for the constraints on type, behavior, and structure are propagated along the proposed mechanism paths of each hypothesis. Three forms of constraint propagation are employed in the simulation of causal graphs, corresponding to three types of values: qualitative regions, ranges of orders of magnitude, and the subjects and objects of relations.

4.1.2.1 Qualitative Calculi

Values for the sign, direction, alignment, bias, and displacement constraints are represented by qualitative regions. Table C.1 through Table C.5 in Appendix C contain qualitative calculi which show how values for these constraints are propagated along mechanism paths.

A non-zero sign {*Negative Positive*}, when propagated across a mechanism which imposes an inverse dependence between cause and effect: *Negative*, is still non-zero: {*Positive Negative*}.

Signs are also affected by any change in orientation between the quantity at the cause and the quantity at the effect. Table C.6 contains a qualitative calculus for combining signs and orientation changes. The observed difference in sign between two events whose signs are *Positive* but whose relative orientation is *Skewed* is the ambiguous {*Negative Zero Positive*}.

The direction {*Parallel Opposite*}, when propagated across a mechanism which imposes an arbitrary deflection: {*Parallel Opposite Perpendicular Skewed*}, becomes maximally ambiguous: {*Parallel Opposite Perpendicular Skewed*}.

An alignment of {*Equal Greater*}, when propagated across a mechanism with the alignment *Less*, which requires the value at the cause to be less than the value at the effect, becomes null: {}. The hypothesis giving rise to this propagation would be no longer viable.

A non-zero sign {*Negative Positive*}, when propagated across a mechanism with a bias towards increase in the effect: {*Down-Up Up-Up*}, becomes unambiguous: {*Positive*}.

The displacement {*Different*}, when propagated across a mechanism which involves a change in site between cause and effect: *Different*, may result in a return to the original location: {*Same Different*}.

4.1.2.2 Range Arithmetic on Orders of Magnitude

Values for the delay and magnitude constraints are represented by ranges

of orders of magnitude. These order of magnitude ranges are propagated according to arithmetic rules, which are given in Appendix D.

Delays are propagated across a mechanism by adding the time lag associated with the mechanism. This time lag is computed by multiplying the distance across the mechanism by the time constant associated with the mechanism. The distance across a hidden, conjectured mechanism—the length of a pipe or wire, for example—cannot be known, except for the trivial case where the physical object at the cause and the physical object at the effect are the same. For each device, a default hidden distance is established. This characteristic distance is taken to be the length of the longest external axis of the device.

For example, suppose the characteristic distance for a device being modelled is $\{\cdot\textsc{range}\cdot\ 2^{-3} : 2^{-3}\}$. For a mechanism with time constant $\{\cdot\textsc{range}\cdot\ 2^{0} : 2^{6}\}$, the computed time lag is $\{\cdot\textsc{range}\cdot\ 2^{-3} : 2^{3}\}$. This time lag, when added to a delay of say, $\{\cdot\textsc{range}\cdot\ 2^{2} : 2^{2}\}$, results in a propagated delay of $\{\cdot\textsc{range}\cdot\ 2^{2} : 2^{3}\}$.

Magnitudes are propagated across a mechanism by multiplying by the efficiency associated with the mechanism. For example, a value for magnitude of $\{\cdot\textsc{range}\cdot\ 2^{1} : 2^{4}\}$, when propagated across a mechanism with efficiency $\{\cdot\textsc{range}\cdot\ 2^{-2} : 2^{0}\}$, is propagated as $\{\cdot\textsc{range}\cdot\ 2^{-1} : 2^{4}\}$.

4.1.2.3 Relations

Values for the type and medium constraints are the subjects and objects of relations. These values are propagated by following chains of relations.

The type conservations and transformations associated with mechanisms are represented by relations whose subjects and objects are quantity types and whose relations are the names of the mechanisms themselves. These relations are shown in Appendix E.

As an example, the quantity type {·TYPE· *Temperature Rate*} is propagated across the mechanism *Thermal-Expansion*, as {·TYPE· *Position Rate*}. Note that the mechanism *Integration* converts any type {·TYPE· *?type Rate*} into {·TYPE· *?type Amount*}. The role of temporal integration in causal modelling is discussed in Section 4.2.

Values for the medium constraint are physical objects which participate in structural relations. A structural relation, or medium, is associated with each mechanism. The propagation of a set of physical objects across a mechanism is the set of physical objects which participate with the given set of physical objects in the type of structural relation associated with the mechanism.

For example, the set of physical objects {*Tire*}, when propagated across a mechanism whose medium is *Joined-To*, is the set of physical objects {*Cylinder*}, providing the relation {·RELATION· *Tire Joined-To Cylinder*} has been asserted, and no other relations {·RELATION· *Tire Joined-To* *} have been asserted.

Structural relations must not be *False* in the interval during which a mechanism is hypothesized to be active. This interval is computed from the value propagated for the delay constraint. The beginning of the interval is the low end of the delay range before propagation across the mechanism; the end of the interval is the high end of the delay range after propagation across the mechanism. Both of these moments are measured from the time of the event taken to be the cause. For example, if the time of the cause is 60, and the value for delay is {·RANGE· $2^2 : 2^3$} before propagation and {·RANGE· $2^3 : 2^4$} after propagation, then the interval during which the medium must be established is {·INTERVAL· 64 : 76}. This calculation involves a conversion from orders of magnitude back to a linear time scale.

For the most part, structural connections and physical objects are hidden inside the "black box" of a device. When no asserted relations are found, a hidden structural relation and physical object are conjectured, as in {·RELATION· *Tire Joined-To ?physical-object-1*}. The conjectured physical object is represented by a variable which may become bound subsequently.

4.1.3 Comparison Rules

The target values derived from the externally observable device events, are compared to the values propagated along the conjectured mechanism paths. The propagated values amount to predictions concerning expected events. The means of comparing prediction and observation differ for the various constraints.

The propagation of type always results in a single value. This prediction must match the observation exactly. The propagation of displacement, sign, direction, alignment, and bias result in a set of qualitative values. The observation must be a member of this set.

Ranges of orders of magnitude are propagated for the delay and magnitude constraints. The observation must intersect the prediction. The test for intersection is:

$high$(prediction) $\geq$ low(observation) or
low(prediction) $\leq$ $high$(observation)

A set of physical objects is propagated for the medium constraint. The observation must be a member of this prediction set. However, there may be an unbound physical object in the propagation set. If this is the case,

the unbound physical object is considered a match and takes on the binding of the observation. Any other unbound physical objects which participate in *Same* relations with the newly-bound physical object also take on this binding. This procedure deals with the problem of equality for conjectured objects in hypothetical worlds [McAllester 80].

We now have a formal basis for deciding when a mechanism hypothesis "explains" or "accounts for" or "is consistent with" an observation of a device. For each of the nine constraints on type, behavior, and structure, the values propagated must match observed values according to the specific tests outlined in the preceding paragraphs.

4.2 Temporal Integration

Causal modelling often involves reasoning about values of quantities changing over time. Quantities may reach critical values—thresholds—which result in abrupt changes in behavior. For example, a sustained motion may result in the closing of a valve which arrests fluid flow, or in the loading of a spring which provides a restoring force.

Temporal integration is the means whereby questions such as "What will be the next value of this quantity?" and "For how long will this quantity change?" are answered. The temporal integration procedure employed in the program JACK utilizes the magnitude of the rate, the duration of change, the direction of change, and the value spaces of quantities.

Temporal integration is treated as a special mechanism which may appear in a causal graph like any other mechanism but for which the propagation rules for the constraints on type, behavior, and structure are in some cases different.

Quantity types are propagated across the temporal integration mechanism in a straightforward manner. A seed type of {·TYPE· *?type Rate*} is propagated as {·TYPE· *?type Amount*}.

The delay associated with temporal integration is the interval during which the rate of a changing quantity remains non-zero, until a new stable value for the amount of the quantity is achieved.

This delay can be computed from the familiar relation: $\Delta t = \Delta a / r$ where t is time, a is the amount and r the rate of a quantity.

In causal modelling, the events at which the values of quantities change are often hidden. Delays due to temporal integration rarely can be computed from observed events. Nevertheless, an upper bound on the delay can be computed from the direction of change, the magnitude of the rate, and knowledge about limiting values in the value spaces of quantities.

The direction of change is the value for the bias constraint propagated thus far. The magnitude of the rate is the value for the magnitude constraint propagated thus far. Limiting values for quantities are taken from default value spaces associated with mechanisms. From this information delay due to temporal integration is propagated as shown in Procedure 4.3.

Given: direction of change, magnitude of rate and
value space for amount
a_i = observed value or default limiting value for amount
opposite to direction of change
a_f = observed value or default limiting value for amount
in direction of change
$\Delta a = a_f - a_i$
r = magnitude of rate
$delay = \Delta a \ / \ r$

Procedure 4.3. Delay across integration.

For example, a default initial value of $\{\cdot\textbf{RANGE}\cdot\ 2^{-\infty} : 2^{-\infty}\}$, a default final value of $\{\cdot\textbf{RANGE}\cdot\ 2^{-\infty} : 2^{7}\}$, and a rate of magnitude $\{\cdot\textbf{RANGE}\cdot\ 2^{0} : 2^{3}\}$ results in a propagated delay of $\{\cdot\textbf{RANGE}\cdot\ 2^{-\infty} : 2^{7}\}$.

The sign of the new value achieved after temporal integration is constrained by the delay, the direction of change, the magnitude of the rate, and limiting values in the value spaces of quantities. Signs are propagated across the temporal integration mechanism as shown in Procedure 4.4.

Given: delay, direction of change, magnitude of rate and
value space for amount
Δt = delay
r = magnitude of rate
a_i = observed value or default limiting value for amount
opposite to direction of change
a_f = observed value or $a_i + \Delta t \cdot r$ or
default limiting value for amount in direction of change
s_i = sign of a_i
s_f = sign of a_f
$sign$ = all signs, inclusive, from s_i to s_f in $\{Negative\ Zero\ Positive\}$

Procedure 4.4. Sign across integration.

For example, an initial sign of *Positive* and a direction of change of *Positive* results in a propagated sign of {*Positive*}. On the other hand, an initial sign of *Positive* and a direction of change of *Negative* results in a propagated sign of {*Negative Zero Positive*}, providing the final value reached has sign *Negative*.

The sign of a value is computed by comparing the position of the value in the value space of the quantity to the position of the zero value in the value space of the quantity. A zero value is always identified in the default value spaces specified in mechanisms.

The amount and rate of a quantity always are presumed to have the same orientation. The deflection due to temporal integration can be inferred from the direction of change (set of signs for the rate) and the resulting set of signs for the amount. If any pair of signs, one from each set, are the same, the deflection includes *Parallel*. If any pair of signs are opposite, the deflection includes *Opposite*. This deflection is propagated according to the qualitative calculus in Table C.2.

The magnitude of the new value achieved via temporal integration is inferred, once again, from the delay, the direction of change, the magnitude of the rate, and limiting values in the value spaces of quantities. Magnitude is propagated across the temporal integration mechanism as shown in Procedure 4.5.

Given: delay, direction of change, magnitude of rate and
value space for amount
Δt = delay
r = magnitude of rate
a_i = observed value or default limiting value for amount
opposite to direction of change
a_f = observed value or $a_i + \Delta t \cdot r$ or
default limiting value for amount in direction of change
if sign of a_i = sign of a_f
then *magnitude* = $\{\cdot\text{RANGE}\cdot\ 2^{min(low(a_i),low(a_f))} : 2^{max(high(a_i),high(a_f))}\}$
else *magnitude* = $\{\cdot\text{RANGE}\cdot\ 2^{-\infty} : 2^{max(high(a_i),high(a_f))}\}$

Procedure 4.5. Magnitude across integration.

The functions *low* and *high* return, respectively, the low and high order of magnitude in a range of orders of magnitude.

The special zero value $2^{-\infty}$ is taken to be the smallest possible resulting magnitude whenever temporal integration may result in a zero crossing.

For example, an initial value of 40 and a final value of -10 results in the propagated magnitude $\{\cdot\textsc{range}\cdot\ 2^{-\infty} : 2^{5}\}$.

Inequalities between values of the amount and rate of the same quantity are not well-defined. The alignment imposed by the temporal integration mechanism is taken to be the maximally ambiguous {*Less Equal Greater*}. This alignment is propagated according to the qualitative calculus in Table C.3.

Values for the bias constraint are propagated intact across the temporal integration mechanism so that the direction of approach to a threshold value is preserved. This information can contribute to the determination of the admissibility of hypotheses. For example, an enablement hypothesis involving, say, a *Pneumatic-Valve* mechanism must be associated with an increase towards positive values; a disablement hypothesis involving the same mechanism must be associated with a decrease towards the zero value.

The amount and rate of a quantity are always associated with the same physical object. The contribution of the temporal integration mechanism for the displacement constraint is the identity displacement *Same*. Values for displacement are propagated as for any other mechanism, according to the qualitative calculus in Table C.5.

Similarly, the medium for temporal integration is the identity structural relation *Same* and this constraint is propagated as for any other mechanism, by extending chains of structural relations.

4.3 Handling One Exponent—Linear Mechanism Paths

One of the sources of complexity in the causal modelling problem is due to uncertainty about the lengths of the mechanism paths within a device. Recall that the number of possible mechanism paths between an arbitrary pair of device events is $\mathcal{O}(m^l)$ where m is the number of possible mechanisms and l is the length of the path.

The search for device models always begins at the root of the ordering on hypotheses—the linear mechanism path level. This part of the causal modelling procedure is shown in Procedure 4.6.

Given: a timeline, a mechanism vocabulary,
  a maximum mechanism path length l_{max},
For each pair of events from the timeline
  consisting of one cause and one effect,
  with the cause not taking place after the effect
  Generate linear mechanism path hypotheses
    of up to length l_{max}

Procedure 4.6. Causal modelling procedure up to linear mechanism paths.

4.4 Handling the Other Exponent—Mechanism Interactions

The other source of complexity in the causal modelling problem is due to uncertainty about interactions among mechanisms. Recall that the worst-case number of possible hypotheses becomes $\mathcal{O}(m^{lp})$ when interactions among mechanism paths are considered, where p is the number of interactions and m^l is the number of hypotheses associated with any mechanism path.

Three types of mechanism interaction are distinguished: *enablement*, where one mechanism arranges for the preconditions of another mechanism to become satisfied; *disablement*, where one mechanism arranges for the preconditions of another mechanism to become unsatisfied; and *equilibrium*, where the contributions of separate mechanisms come into balance.

An example of enablement is a switch being closed and permitting the flow of electricity. An example of disablement is a latch being engaged and arresting a motion. An example of equilibrium is the steady level of water in a sink when the flow in at the faucet balances the flow out at the drain.

Mechanism interaction hypotheses appear below linear mechanism path hypotheses in the ordering on hypotheses. However, the causal modelling system does not extend all hypotheses generated at the linear mechanism path level. In the interest of keeping the hypothesis set manageably small at all times, a set of heuristics is employed for deciding whether or not to consider hypotheses involving mechanism interactions.

All of the heuristic rules for justifying level jumps are based on the principle that too-simple hypotheses exhibit characteristic deficiencies. Heuristic rules capture these signatures and justify attempts to repair recognizably deficient hypotheses by extending them into more complex or less constrained hypothesis forms in the ordering on hypothesis types.

The modified causal modelling procedure which includes the heuristically justified level jump to mechanism interactions is shown in Procedure 4.7.

Given: a timeline, a mechanism vocabulary,
  a maximum mechanism path length l_{max},
  and a maximum number of interactions p_{max}
For each pair of events from the timeline
  consisting of one cause and one effect,
  with the cause not taking place after the effect
  $p = 1$
  Generate linear mechanism path hypotheses
    of up to length l_{max}
  Extend: For each hypothesis
    When $p < p_{max}$ and a mechanism interaction heuristic is satisfied
      $p = p + 1$
      Generate mechanism interaction hypotheses
      Go to Extend:

Procedure 4.7. Causal modelling procedure up to mechanism interactions.

4.4.1 Heuristics for Mechanism Interactions

Enablements are characterized by unexplained delays. Once a pending mechanism becomes enabled however, the resulting effect is always as expected. The only exception is a possible decrease in magnitude as in the case of say, a half-open valve. The heuristic for recognizing enablement situations is shown in Procedure 4.8.

Given: a hypothesis.
Either exactly the delay constraint is violated
  or exactly the delay and magnitude constraints are violated

Procedure 4.8. Enablement heuristic.

The signature for disablements is an unexpected zero value occurring after a non-zero effect is expected. The heuristic for recognizing disablement situations is shown in Procedure 4.9.

Given: a hypothesis.
Exactly the delay, sign, magnitude, and bias constraints are violated
and the value of the effect is zero
and the effect is not at a limiting value

Procedure 4.9. Disablement heuristic.

Equilibria also are characterized by an unexpected zero value when the expected effect is non-zero. The zero value may occur after the expected time of occurrence of a non-zero effect. The heuristic for recognizing equilibrium situations is shown in Procedure 4.10.

Given: a hypothesis.
Either exactly the sign, magnitude, and bias constraints are violated
or exactly the delay, sign, magnitude, and bias constraints are violated
and the value of the effect is zero
and the effect is not at a limiting value

Procedure 4.10. Equilibrium heuristic.

Note that the disablement and equilibrium heuristics are indistinguishable when there is an unexplained delay.

4.4.2 Combination Rules for Enablement and Disablement Hypotheses

Once the possibility of an enablement or disablement situation is established, causal modelling shifts from hypothesizing linear mechanism paths to hypothesizing interacting mechanism paths.

Candidate events for the initial cause of an enabling or disabling mechanism path are those events which are strictly before the effect event. Enablements and disablements always involve integrating a quantity over a non-zero temporal interval—a switch is closed, a pressure is raised, etc. Enablements and disablements do not occur instantaneously.

An interaction may take place at any point along a mechanism path. For example, a hydraulically induced motion may be inhibited by closing a valve on fluid flow or by directly latching a piston.

Procedure 4.11 is used to generate enablement and disablement interaction hypotheses.

Given: a linear mechanism path hypothesis
For each event from the timeline before the effect event
  For each mechanism in the mechanism path
    Hypothesize an enablement/disablement
      interaction at the given mechanism

Procedure 4.11. Enablement and disablement hypothesis generation.

After the separate contributions of interacting mechanism paths are combined at points of interaction, values for each of the constraints are propagated further along the remainder of the original mechanism path according to the propagation rules for linear mechanism paths. See Figure 4.1.

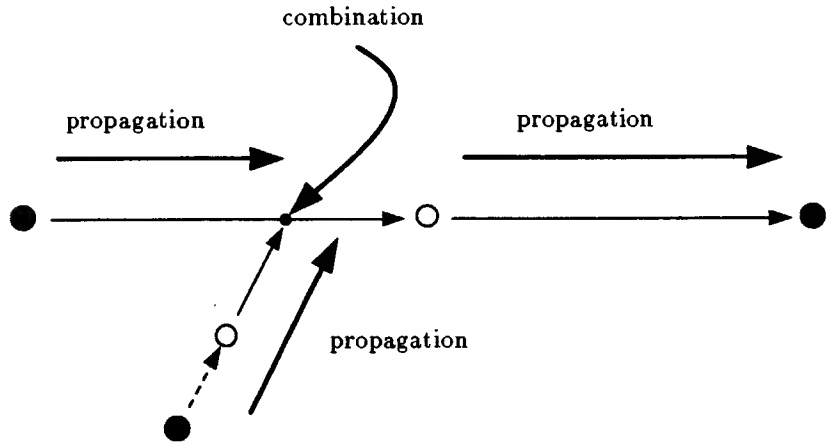


Figure 4.1. Propagation and combination in mechanism interaction hypotheses.

All that remains to be described is how values for the constraints on type, behavior, and structure are combined at interaction points under enablement and disablement.

The types propagated along the interacting mechanism paths must match exactly. For example, a *Vent*, which controls changes in *Temperature*, cannot interact with *Electricity* which involves *Charge* quantities.

Delay is measured from the time of interaction, as described in Procedure 4.12. See also Figure 4.2.

Given: moment of primary cause t_{c_1},
moment of interacting cause t_{c_2}
delay along primary path $delay_{c_1}$,
and delay along interacting path $delay_{c_2}$
$delay = max(delay_{c_1}, delay_{c_2} - (t_{c_1} - t_{c_2})$

Procedure 4.12. Delay for mechanism interactions.

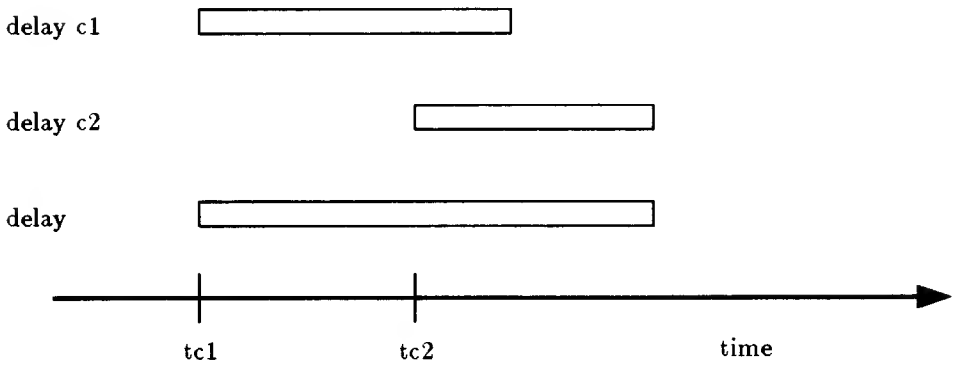


Figure 4.2. Delay for mechanism interactions.

Values for sign are composed multiplicatively, with values along the enablement or disablement path being restricted to {*Zero Positive*}. A fully enabled mechanism corresponds to multiplying by +1; a fully disabled mechanism corresponds to multiplying by 0. The qualitative calculus for combining signs under multiplication is the same used for propagating signs along linear mechanism paths, and appears in Table C.1.

Direction is entirely unaffected by enablement or disablement: values for direction are passed directly.

Values for magnitude also are composed multiplicatively, with values in the value space of the quantity along the enablement or disablement path being normalized to the order of magnitude range {·RANGE· $2^{-\infty} : 2^0$} (corresponding to the range [0 : 1] on a linear scale).

Alignment also is unaffected by enablement or disablement; as long as the alignment propagated along the enablement or disablement path is non-null, the alignment propagated along the primary causal path is passed directly.

Values for bias are combined in the same manner as values for sign.

Displacements are mostly passed unaffected; the exception is when the value along both interacting paths is {*Same*} and the physical objects associated with the initial cause event on the each path are not the same. In this case, the null set is passed on to mark the contradiction.

Similarly, the physical objects propagated for the medium constraint along the enablement or disablement path and the primary causal path must match if they are bound; otherwise they are asserted to be the same physical object.

4.4.3 Combination Rules for Equilibrium Hypotheses

The procedure for generating equilibrium hypotheses is nearly the same as the procedure for generating enablement and disablement hypotheses. The only difference is that events occurring simultaneously with the effect event also are candidate initial causes for the interacting mechanism path. Equilibria may be achieved instantaneously whereas enablements and disablements always involve a threshold value being reached over a non-zero temporal interval. This procedure is shown in Procedure 4.13.

```
Given: a linear mechanism path hypothesis
For each event not after the effect event
   For each mechanism in the mechanism path
      Hypothesize an equilibrium interaction
         at the given mechanism
```

Procedure 4.13. Equilibrium hypothesis generation.

Some of the rules for combining values for the constraints on type, behavior, and structure for equilibrium hypotheses are different from the combination rules for enablement and disablement hypotheses.

Enablement and disablement situations may be thought of as mechanism *conjunction*; all preconditions must be satisfied or all enabling mechanisms must be active for an effect to occur. Not surprisingly, many of the rules for composition are multiplication rules.

On the other hand, equilibrium situations are instances of mechanism *disjunction*; the contributions of the interacting mechanisms are separable.

Some effect occurs whether or not there is an interaction. Correspondingly, many of the rules for composition are addition rules.

For example, values for sign are combined additively rather than multiplicatively under equilibrium. The qualitative calculus for composing signs under addition appears in Table C.7.

Values for direction also are combined additively. The calculus employed for this purpose may be thought of as qualitative vector addition. This calculus is found in Table C.8.

Values for magnitude are composed according to the addition rule for ranges of orders of magnitude.

Biases are composed under equilibrium as are signs, using the qualitative calculus for sign addition in Table C.7.

The combination rules for the type, delay, alignment, displacement, and medium constraints are the same as for enablement and disablement hypotheses.

4.5 Handling Lost Constraint—Hidden Inputs

Implicit in the construction of linear mechanism path and mechanism interaction hypotheses is the assumption that the initial cause or primitive input on each conjectured mechanism path is always among the observable events of a device. This assumption precludes, for example, hypothesizing an unknown electrical source such as a hidden battery.

The cost of removing the assumption of no hidden inputs is not a greater number of possible hypotheses in the worst case, but a sharply reduced capability for testing hypotheses. Without observations against which to compare the results of propagating values for the constraints on type, behavior, and structure, hypothesizing becomes a case of "almost anything goes." There must be compelling reasons for entertaining hidden input hypotheses.

The modified causal modelling procedure which includes the heuristically justified level jump to hidden inputs is shown in Procedure 4.14.

Given: a timeline, a mechanism vocabulary,
  a maximum mechanism path length l_{max},
  and a maximum number of interactions p_{max}
For each pair of events from the timeline
  consisting of one cause and one effect,
  with the cause not taking place after the effect
  $p = 1$
  Generate linear mechanism path hypotheses
    of up to length l_{max}
  Extend: For each hypothesis
    When $p < p_{max}$ and a mechanism interaction heuristic is satisfied
      $p = p + 1$
      Generate mechanism interaction hypotheses
    When $p < p_{max}$ and the hidden input heuristic is satisfied
      $p = p + 1$
      Generate hidden input hypotheses
      Go to Extend:

Procedure 4.14. Causal modelling procedure up to hidden inputs.

4.5.1 Heuristic for Hidden Inputs

Some mechanisms can participate only in enablement and disablement interactions; they cannot stand alone along linear mechanism paths. These mechanisms are enumerated in Appendix B. For example, an explanation for a fluid flow in terms of an enabling opened valve is incomplete; there also must be a fluid source. Similarly, an explanation for cooling in terms of a pressure decrease enabling evaporation also is incomplete; there also must be a heat sink. The appearance of an "interaction-only" mechanism on a linear mechanism path implies a missing input. The heuristic for recognizing hidden input situations is shown in Procedure 4.15.

Given: a linear mechanism path or mechanism interaction hypothesis.
There is an enabling or disabling mechanism
  on a linear mechanism path

Procedure 4.15. Hidden input heuristic.

4.5.2 Propagation and Combination Rules for Hidden Input Hypotheses

The construction of hidden input hypotheses involves extending linear mechanism path hypotheses into interaction hypotheses at the mechanisms which must participate in interactions. This process involves inverse propagation and combination of constraint values and is described in Procedure 4.16. See also Figure 4.3.

Given: a candidate hypothesis for a hidden input
Propagate forward from cause event to point of interaction
Propagate backward from effect event to point of interaction
Invert combination at point of interaction
Propagate backward along new mechanism path from point of interaction

Procedure 4.16. Hidden input hypothesis generation.

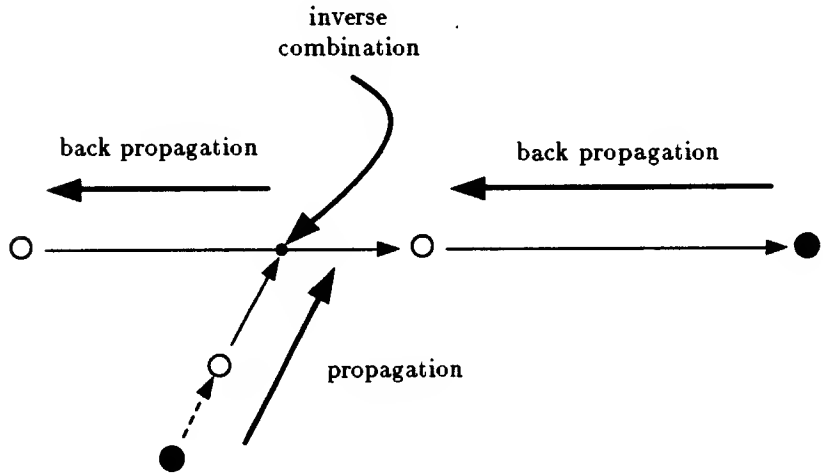


Figure 4.3. Propagation and combination in hidden input hypothesis generation.

The inverse propagation rules for the constraints on type, behavior, and structure are defined as follows:

The qualitative calculi for the sign, direction, alignment, bias, and displacement constraints are traversed in the backward direction.

The appropriate inverse arithmetic rules for ranges of orders of magnitude are substituted for the delay and magnitude constraints: subtraction for addition and division for multiplication. In particular, delay is propagated by subtracting the time lag associated with a mechanism from the time of an effect; magnitude is propagated by dividing the magnitude of an effect by the efficiency of a mechanism.

The relations for the type and medium constraints are traversed from object to subject, rather then from subject to object.

The propagation rules for temporal differentiation are based on the formula $r = \Delta a/\Delta t$. This formula is used to infer the interval during which the rate is non-zero for the delay constraint, the direction of change for the sign constraint, and the magnitude of the rate for the magnitude constraint. Δa is bounded by the current value for the amount, given by the value for the magnitude constraint propagated thus far, and default limiting values in the value spaces of quantities. Δt is bounded by the observed time of the effect and the earliest moment on the timeline. r is not bounded by default limits.

The temporal differentiation rules for the direction, alignment, bias, displacement, type, and medium constraints are the same as those for temporal integration.

The inverse combination rules at points of interaction for the constraints on type, behavior, and structure are defined as follows:

The delay between the time of the contribution due to the hidden input and the time of the effect is bounded by the delay propagated backward from the effect and the beginning of the timeline.

When an equilibrium interaction is being inverted, values for sign are combined by traversing the additive calculus for sign in the backward direction. When an enablement or disablement interaction is being inverted, values for sign are combined by traversing the multiplicative calculus for sign in the backward direction, with the contribution of the enabling or disabling path being restricted to $\{Zero\ Positive\}$.

When an equilibrium interaction is being inverted, values for magnitude are combined with the subtraction rule for ranges of orders of magnitude. When an enablement or disablement interaction is being inverted, values for magnitude are combined with the division rule for ranges of orders of magnitude, with the contribution of the enabling or disabling path being normalized to the range $[0 : 1]$.

The inverse combination rule for the bias constraint is the same as the inverse combination rule for the sign constraint.

The inverse combination rules for the type, direction, alignment, displacement, and medium constraints are the same as the corresponding direct combination rules.

4.6 Handling Sources and Sinks—Cycles

Potential sources and sinks in a physical system sometimes can be avoided by balancing one against the other. Such a balance can be achieved in a cycle where gains alternate with losses so that there is never an unbounded increase or decrease. An example of such a synergistic cycle is the circulation of refrigerant in a refrigerator: a single material is alternately evaporated, taking up heat from the interior, and condensed, giving off heat to the environment. The net heat gain or heat loss in this material remains always bounded.

Synergistic cycles, in which potential sources and sinks are removed, are to be contrasted with iterative cycles, in which an increase or decrease is built to a threshold. Only the synergistic type of cycle is being treated here. Furthermore, only hidden input hypotheses serve as candidates for extension into this kind of cycle hypothesis. Synergistic cycle hypotheses show how the conjectured sources and sinks of hidden input hypotheses can be avoided. (Inputs can be either sources or sinks—the air entering a tire gauge and the air leaving a vacuum cleaner both serve as inputs). The same cycle construction exercise is pointless for linear mechanism path or mechanism interaction hypotheses, whose initial causes always are among the observable, declared inputs of a device.

The modified causal modelling procedure which includes the heuristically justified level jump to cycles is shown in Procedure 4.17.

Given: a timeline, a mechanism vocabulary,
  a maximum mechanism path length l_{max},
  and a maximum number of interactions p_{max}
For each pair of events from the timeline
  consisting of one cause and one effect,
  with the cause not taking place after the effect
  $p = 1$
  Generate linear mechanism path hypotheses
    of up to length l_{max}
  Extend: For each hypothesis
    When $p < p_{max}$ and a mechanism interaction heuristic is satisfied
      $p = p + 1$
      Generate mechanism interaction hypotheses
    When $p < p_{max}$ and the hidden input heuristic is satisfied
      $p = p + 1$
      Generate hidden input hypotheses
      Go to Extend:
For each pair of hidden input hypotheses
  When the cycle heuristic is satisfied
    Generate a cycle hypothesis

Procedure 4.17. Causal modelling procedure up to cycles.

4.6.1 Heuristic for Cycles

Synergistic cycles are characterized by a potential source together with a potential sink. The heuristic for recognizing cycle situations is shown in Procedure 4.18.

Given: two hidden input hypotheses.
s_1 = the set of signs associated with
the hidden input of one hypothesis
s_2 = the set of signs associated with
the hidden input of the other hypothesis
There is a pair of signs,
one from s_1 and one from s_2,
which are of opposite value

Procedure 4.18. Cycle heuristic.

4.6.2 Combination Rules for Cycle Hypotheses

A cycle hypothesis is constructed from a pair of hidden input hypotheses by additively combining the hidden inputs from the two hypotheses—one source and one sink. This procedure is shown in Procedure 4.19. See also Figure 4.4.

Given: two hidden input hypotheses.
Hypothesize a cycle interaction between
the two hidden input event nodes,
one from each hypothesis.

Procedure 4.19. Cycle hypothesis generation.

The rules for additive combination in cycle hypotheses are nearly the same as those for constructing equilibrium hypotheses.

Values for the type constraint must match exactly.

Propagated delays must be "out-of-phase" because gains and losses alternate in a synergistic cycle. Given that values for the delay constraint are propagated as ranges, this requirement is easy to satisfy. Only equal point ranges fail to do so.

The *Zero* value must result from the combination of values for the sign and bias constraints.

The *Opposite* value must result from the combination of values for the direction constraint.

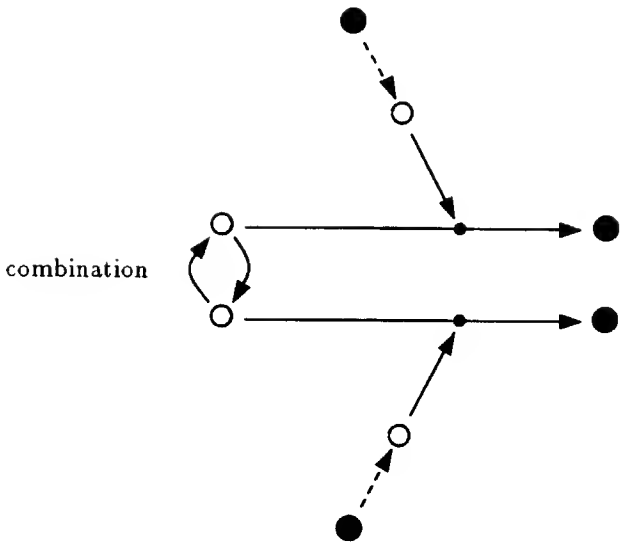


Figure 4.4. Combination in cycle hypotheses.

The ranges of orders of magnitude propagated for the magnitude constraint must overlap so that upon combination, the low end of the resulting range is the zero value $2^{-\infty}$.

The values propagated for the alignment constraint must be non-null.

The physical object associated with the two hidden inputs must be the same physical object. Otherwise the gain from one and the loss from the other do not offset.

Values for the displacement constraint violate this requirement when {*Same*} is propagated for both hypotheses and the physical objects associated with the final effect in each hypothesis are not the same.

Similarly, the two physical objects propagated for the medium constraint must be the same physical object, or both must be unbound, at which point they are asserted to be the same physical object.

4.7 A Detailed Example

In this section, I work through a detailed example of hypothesis construction. The hypothesis is shown in Figure 4.5. This example serves to illustrate the

propagation and combination rules and the temporal integration procedures which make up the causal and qualitative simulation method, the comparison rules for verifying predicted events against observed events, and the heuristics for traversing the hypothesis ordering.

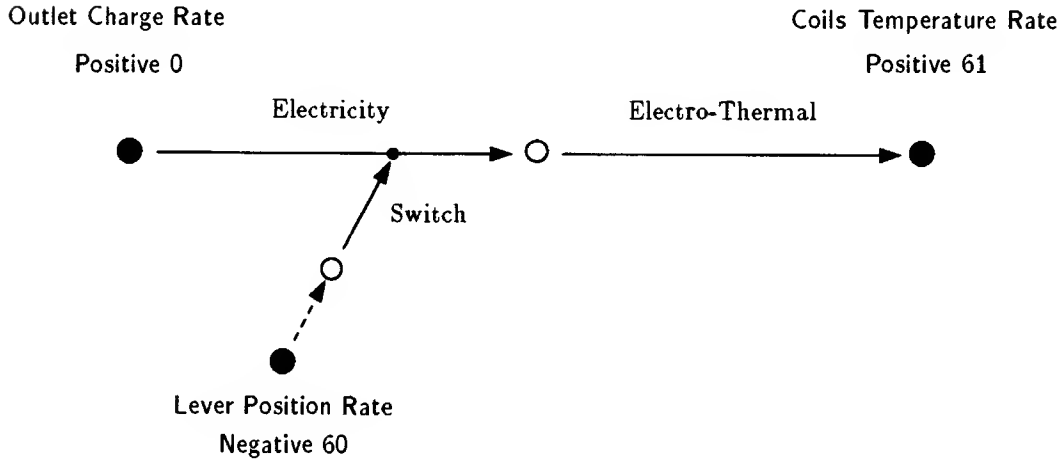


Figure 4.5. An example hypothesis.

In this example, the device event {·**EVENT**· *Outlet Charge Rate Positive 0*} is taken to be the cause and the device event {·**EVENT**· *Coils Temperature Rate Positive 61*} is taken to be the effect. One of the generated hypotheses is the linear mechanism path {·**MECHANISM PATH**· *Electricity Electro-Thermal*}. The seed event node computed from the cause event is:

{·**EVENT NODE**·
:type {·**TYPE**· *Charge Rate*}
:delay {·**RANGE**· $2^{-\infty} : 2^{-\infty}$}
:sign *Positive*
:direction *Parallel*
:magnitude {·**RANGE**· $2^{3} : 2^{3}$}
:alignment {*Less Equal Greater*}
:bias *Positive*
:displacement *Same*
:medium *Outlet*}

The target event node computed from the cause event and the effect event is:

{·EVENT NODE·
 :type {·TYPE· *Temperature Rate*}
 :delay {·RANGE· $2^{6} : 2^{6}$}
 :sign *Positive*
 :direction *Parallel*
 :magnitude {·RANGE· $2^{1} : 2^{1}$}
 :alignment {*Less Equal Greater*}
 :bias *Positive*
 :displacement *Different*
 :medium *Coils*}

The event node which represents the effect of the *Electricity* mechanism is computed via the propagation rules. This event node is:

{·EVENT NODE·
 :type {·TYPE· *Charge Rate*}
 :delay {·RANGE· $2^{-\infty} : 2^{-\infty}$}
 :sign {*Positive*}
 :direction {*Parallel Opposite Perpendicular Skewed*}
 :magnitude {·RANGE· $2^{-4} : 2^{10}$}
 :alignment {*Less Equal Greater*}
 :bias {*Positive*}
 :displacement {*Different*}
 :medium {*Coils*}}

The event node which represents the effect of the *Electro-Thermal* mechanism, computed similarly, is:

{·EVENT NODE·
 :type {·TYPE· *Temperature Rate*}
 :delay {·RANGE· $2^{-\infty} : 2^{-\infty}$}
 :sign {*Positive*}
 :direction {*Parallel Opposite Perpendicular Skewed*}
 :magnitude {·RANGE· $2^{-11} : 2^{17}$}
 :alignment {*Less Equal Greater*}
 :bias {*Positive*}
 :displacement {*Different*}
 :medium {*Coils*}}

This event node is incompatible with the target event node. In particular, this hypothesis fails because the delay constraint is violated. However,

this partial failure triggers the enablement interaction recognition rule. The hallmark of an enablement interaction is an unexplained delay.

Next the device event {·EVENT· *Lever Position Rate Negative 60*} is taken to be an additional cause. One of the generated enablement path hypotheses {·MECHANISM PATH· *Integration Switch*}. The seed event computed from this second cause is:

{·EVENT NODE·
 :type {·TYPE· *Position Rate*}
 :delay {·RANGE· $2^{-\infty} : 2^{-\infty}$}
 :sign *Negative*
 :direction *Parallel*
 :magnitude {·RANGE· $2^{-3} : 2^{-3}$}
 :alignment {*Less Equal Greater*}
 :bias *Negative*
 :displacement *Same*
 :medium *Lever*}

The event node which represents the effect of the *Integration* mechanism—computed via the temporal integration procedures—is:

{·EVENT NODE·
 :type {·TYPE· *Position Amount*}
 :delay {·RANGE· $2^{0} : 2^{0}$}
 :sign {*Negative*}
 :direction {*Parallel*}
 :magnitude {·RANGE· $2^{-3} : 2^{-3}$}
 :alignment {*Less Equal Greater*}
 :bias {*Negative*}
 :displacement {*Same*}
 :medium {*Lever*}}

The event node propagated past the *Switch* mechanism is:

{·EVENT NODE·
:type {·TYPE· *Charge Rate*}
:delay {·RANGE· $2^0 : 2^0$}
:sign {*Negative*}
:direction {*Parallel Opposite Perpendicular Skewed*}
:magnitude {·RANGE· $2^{-3} : 2^{-3}$}
:alignment {*Less Equal Greater*}
:bias {*Negative*}
:displacement {*Different*}
:medium {*Coils*}}

This event is composed with the event node propagated past the *Electricity* mechanism via the combination rules for enablement interactions. The result is:

{·EVENT NODE·
:type {·TYPE· *Charge Rate*}
:delay {·RANGE· $2^6 : 2^6$}
:sign {*Positive*}
:direction {*Parallel Opposite Perpendicular Skewed*}
:magnitude {·RANGE· $2^{-4} : 2^{10}$}
:alignment {*Less Equal Greater*}
:bias {*Positive*}
:displacement {*Different*}
:medium {*Coils*}}

The result of a final propagation past the *Electro-Thermal* mechanism is the event node:

{·EVENT NODE·
:type {·TYPE· *Temperature Rate*}
:delay {·RANGE· $2^6 : 2^6$}
:sign {*Positive*}
:direction {*Parallel Opposite Perpendicular Skewed*}
:magnitude {·RANGE· $2^{-11} : 2^{17}$}
:alignment {*Less Equal Greater*}
:bias {*Positive*}
:displacement {*Different*}
:medium {*Coils*}}

This event node is compatible with the target event node. This hypothesis is admitted.

4.8 Refining Hypotheses

A hypothesis may explain some of the behavior of a device and yet be inconsistent with other behavior. For example, several linkages may be proposed to account for the motion of the slide in a pocket tire gauge: a coupling with attachment, a coupling based on contact only, a ratchet. However, an attachment coupling is inconsistent with the slide remaining where it is when the gauge later is removed from the tire; a ratchet is inconsistent with the slide later being pushed back into the cylinder.

Hypothesis refinement distinguishes theory formation, which operates on as many examples of the behavior of a device as are available, from explanation, which operates on isolated examples of the behavior of a device with no concern for global consistency.

The constraints on type, behavior, and structure are abstractions of physical and causal principles; they capture necessary but insufficient conditions which all causal models must satisfy. Accordingly, false positives are assumed to be possible but false negatives are not. The only form of refinement in the causal modelling process is specialization. Hypotheses which are found to be inconsistent with additional examples of device behavior are retracted.

The procedure for refining hypotheses is given in Procedure 4.20.

For each p-tuple of cause events
Form the set of quantities associated with the p cause events
Collect the hypotheses indexed under this set of cause quantities
For each existing hypothesis
  Propagate constraints in the context of the hypothesis
    to form a prediction
  Search timeline for observation
  When prediction and observation
    then
    When predicted change in delay does not match
      observed change in delay
      then FALSE ; *disproportionate delay*
    When predicted change in magnitude does not match
      observed change in magnitude
      then FALSE ; *disproportionate magnitude*
    else TRUE ; *verified effect*
  When prediction and no observation
    then FALSE ; *unverified effect*
  When no prediction and observation
    then FALSE ; *unexpected effect*
  When no prediction and no observation
    then TRUE ; *no unexpected effect*

Procedure 4.20. Hypothesis Refinement.

Admitted hypotheses are indexed under the set of quantities associated with the initial cause events of the hypothesis. For example, the hypothesis of Figure 4.5 is indexed under {{·QUANTITY· *Outlet Charge Rate*} {·QUANTITY· *Lever Position Rate*}}. Then, as other events in a timeline are processed, existing hypotheses for a given set of cause quantities can be retrieved and used to form predictions.

For each retrieved hypothesis. there is a predicted effect event for the new set of cause events exactly when all of the values propagated for the constraints on type, behavior, and structure are non-null. Global consistency across multiple examples of behavior is ensured by propagating constraints in the context of any assertions made during the generation of the original hypothesis. For example. in a *Contact-Coupling* hypothesis, an assertion is made concerning the relative position of the physical objects associated with cause and effect. This inequality assertion must not be violated by other

observed motions or lack of motions concerning those physical objects.

The type, medium, delay, sign, and magnitude constraints are used to focus the search for an observation from the timeline. The type and medium constraints jointly determine the quantity associated with the expected event. Quantities are uniquely defined by a physical object, provided by the value of the medium constraint, and a type. The value of the delay constraint, after translation from order of magnitude scale to linear scale, determines the temporal interval during which to expect the effect. The sign and magnitude constraints jointly determine the range of values to expect. These are the values from the value space of the predicted quantity which are compatible with the propagated sign and magnitude. The predicted quantity, moment, and value completely determine the expected event to be sought on the timeline.

4.8.1 Linear Mechanism Paths

Forming predictions for linear mechanism paths is straightforward. An effect is expected as long as none of the constraints are violated. The cases are enumerated in Table 4.1.

Mechanism Path	*Effect*
active	expected
inactive	not expected

Table 4.1. Prediction for linear mechanism paths.

4.8.2 Enablement and Disablement Interactions

Forming predictions for enablement and disablement interactions is slightly more complicated. The contributions of the primary path and the enabling or disabling path must be teased apart. The cases are enumerated in Table 4.2 and Table 4.3.

Primary Path	*Enabling Path*	*Effect*
active	active	expected
active	inactive	not expected
inactive	active	not expected
inactive	inactive	not expected

Table 4.2. Prediction for enablement interactions.

An effect is expected only as long as none of the constraints are violated for either the primary path or the enabling path.

Primary Path	*Disabling Path*	*Effect*
active	active	expected
active	inactive	effect due to primary path expected
inactive	active	expected
inactive	inactive	not expected

Table 4.3. Prediction for disablement interactions.

The expected effect for a disablement situation is that the value of a quantity will cease to change. This effect is expected whenever none of the constraints are violated for the disabling path—whether or not the constraints are violated for the primary path. Furthermore, when the primary path is intact but the disabling path is not, a different effect is expected—the effect due to the primary path alone.

4.8.3 Equilibrium Interactions

Forming predictions for equilibrium interactions also requires the contributions of the opposing paths to be teased apart. The cases are enumerated in Table 4.4.

One Path	*Other Path*	*Effect*
active	active	expected
active	inactive	effect due to the one path expected
inactive	active	effect due to the other path expected
inactive	inactive	not expected

Table 4.4. Prediction for equilibrium interactions.

The expected effect for an equilibrium situation also is that the value of a quantity will cease to change. This effect is expected only as long as none of the constraints are violated for either of the opposing paths. If these constraints are violated in exactly one of the opposing paths, the expected effect is the one due to the remaining intact path alone.

4.8.4 Hidden Inputs

Refinement of hidden input hypotheses is pointless. Whatever predic-

tion may be generated—smaller hidden input, earlier hidden input, no hidden input—it is unverifiable because hidden inputs are, by definition, unobservable.

4.8.5 Cycles

Forming predictions for cycles is nearly the same as forming predictions for equilibrium situations. The alternating, additive contributions of the cycle halves have to be teased apart. The cases are enumerated in Table 4.5.

Source Path	*Sink Path*	*Effect*
active	active	expected
active	inactive	effect due to source path expected
inactive	active	effect due to sink path expected
inactive	inactive	not expected

Table 4.5. Prediction for cycles.

However, these predictions are interpreted differently for cycle hypotheses. The expected effect of a synergistic cycle is that a potential increase or decrease in a quantity remains always bounded. This effect is expected only when the constraints are satisfied for both halves of a cycle. Whenever only one of the alternating contributions is active the synergy is compromised and the latent source or sink emerges. A cycle hypothesis is tenable only as long as both or neither of the cycle halves is active for all observations.

Several examples of hypothesis refinement are discussed in Chapter 5.

5. Examples: These Are the Models That JACK Built

In this chapter, I describe the operation of the program JACK on several examples. For each device example, I discuss how much of the complexity of the "real" physical system is examined by the program, which aspects of the causal modelling process are exercised by the example, and how the reasoning exhibited by the program JACK both captures important aspects of and falls short of an engineer's understanding of the device.

As part of this scrutiny of performance, I examine both the "correct" hypothesis generated by the causal modelling system—the one which captures, albeit approximately, the standard design for the device—and some of the "bogus" hypotheses which, after the first chuckle, sometimes turn out to be perfectly plausible. These hypotheses reflect the same physical and causal principles as the target hypothesis and in some cases represent genuine, if abstract, alternate designs for the devices.

5.1 The Toaster

In this example, I test the program JACK on a simplified version of the common household toaster. The observation of the toaster, which may be found in Figure 5.1 runs as follows: Initially, the lever and carriage are in their upright position and motionless. The dial has some particular stable setting. The coils are cold, the bread is white and neither is changing. Both electricity at the outlet and gravity are declared to be available as primitive causes, or device inputs. The first thing that happens is the lever and carriage move downward together; the lever's motion is declared to be a device input. Next the lever and carriage stop moving together and at the same time, the coils begin heating up. Next the bread begins to get darker; this event is declared to be a final effect, or device output. Some time later, the lever and carriage move upward together. Simultaneously, the coils stop heating and the bread stops getting darker. Immediately, the coils begin to cool. The lever and carriage reach their uppermost position and stop moving. Sometime later, the coils reach a stable temperature.

5.1.1 Distinguishing Properties of the Toaster Example

The toaster example was the first implemented and has remained the primary benchmark against which all modifications to the program are tested. The example is particularly rich and is well suited to this role. The toaster contains electrical, mechanical, and thermal mechanisms. In addition, there

	0:00	1:00	1:01	1:06	3:06	3:07	7:30
Lever Position Amount	Up		Down			Up	
Lever Position Rate	Zero	Negative	Zero		Positive	Zero	
Dial Angle Amount	LM						
Dial Angle Rate	Zero						
Carriage Position Amount	Up		Down			Up	
Carriage Position Rate	Zero	Negative	Zero		Positive	Zero	
Coils Temperature Amount	Off				Hot		Off
Coils Temperature Rate	Zero		Positive		Zero	Negative	Zero
Bread Appearance Amount	Untoasted				Golden		
Bread Appearance Rate	Zero			Positive	Zero		
Outlet Charge Amount	On						
Outlet Charge Rate	Positive						
Earth Gravity Amount	G						
Earth Gravity Rate	Zero						

Figure 5.1. Timeline of toaster observation.

are enablements and disablements: switches opening and closing, latches being engaged and disengaged. A second observation of the toaster, in which the bread turns out lighter, affords opportunities to refine hypotheses.

5.1.2 Reasoning About the Toaster

Figure 5.2 shows one of the hypotheses constructed by the program JACK to account for the temperature increase of the coils in the toaster. Let us examine the steps taken by the causal modelling system in arriving at this

hypothesis. First, the program JACK attempts to construct hypotheses consisting solely of linear mechanism paths; this is the simplest type of hypothesis in the ordering of hypothesis types. As it happens, no linear mechanism path hypothesis satisfies all the constraints on type, behavior and structure. However, some of the failed hypotheses trigger the heuristic for suspecting enablement situations: there is an unexplained delay. Among these is the linear mechanism path hypothesis {·MECHANISM PATH· *Electricity Electro-Thermal*}. A delay is observed between the event describing current at the outlet and the event describing the temperature increase at the coils. However, both of these mechanisms have zero delays: electricity propagates at the speed of light and the electro-thermal transformation has no distance to cover, taking place inside a single physical object.

The program JACK now attempts to construct hypotheses describing enablement interactions. Candidates for the enabling event are those events which occur before the effect event, in this case before the temperature change of the coils. The causal modelling system is able to construct a hypothesis involving the motion of the lever which satisfies all the constraints. This is the hypothesis in Figure 5.2. The enabling path in this hypothesis includes an *Integration* mechanism; the switch moves from one position to another as part of the enablement process. This integration episode accounts for the observed delay between cause and effect which the original hypothesis could not explain. For enablement interactions, the composed delay is the maximum of the delay due to the primary causal path and the delay due to the enablement path. The new value for the position of the lever is achieved at the same moment at which the temperature of the coils begins to increase. This observation is consistent with the hypothesis; once the flow of electricity is enabled, there is no additional delay associated with the electro-thermal transformation.

The causal modelling system generates another set of enablement hypotheses involving motion of the carriage rather than motion of the lever. This is entirely reasonable given the similarity between the two events: the lever and carriage move at the same rate and stop moving at the same time. Without knowledge of internal structural connections which might disambiguate whether it is the lever or the carriage which interacts with the flow of electricity, the program JACK has no basis for preferring one over the other.

Another linear mechanism path hypothesis which triggers the enablement heuristic is the hypothesis {·MECHANISM PATH· *Electro-Thermal Conductive-Heat-Flow*}. Here the program JACK proposes that current at the outlet results in a temperature increase at the outlet which is then transferred to the coils via a heat flow. The delay associated with the *Conductive-Heat-Flow* mech-

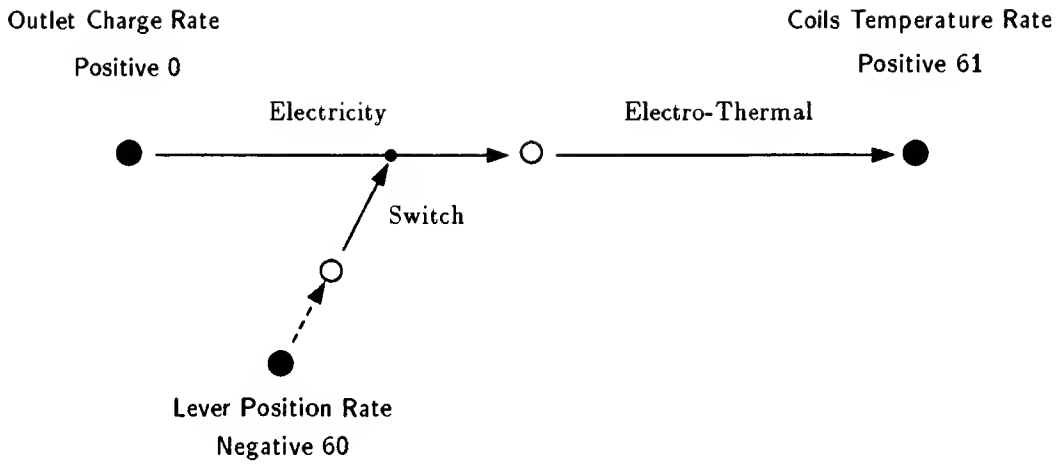


Figure 5.2. Switch hypothesis for coils.

anism is not long enough to account for the observed delay.

The program JACK now attempts to extend this heat flow hypothesis to include an enablement interaction. One of the successful attempts appears in Figure 5.3. Here the motion of the lever is opening a vent which enables heat flow rather than closing a switch which enables electrical flow. The heat flow and vent hypothesis satisfies the delay constraint in the same way as the electricity and switch hypothesis. In fact, the causal modelling system generates a set of heat flow and vent hypotheses similar to the set of electricity and switch hypotheses. The hypotheses involve either the motion of the lever or the motion of the carriage as the initial enabling event and they differ in the inclusion of additional linkages up to the maximum mechanism path length.

The heat flow and vent hypothesis may seem strange because people have enough experience with toasters to know that the outlet does not heat up appreciably and that the cord carries electricity, not heat. However, the hypothesis is physically plausible and is not at variance with the given observation of the toaster. It is instructive at this point to examine the reasoning by which the program JACK prunes out other hypotheses which, at first glance, seem no more strange than the heat flow and vent hypothesis.

In several pruned hypotheses, gravity is proposed as the initial enabling event which produces the motion which closes a switch or opens a vent. Such hypotheses seem perfectly reasonable, but it turns out that they are eliminated

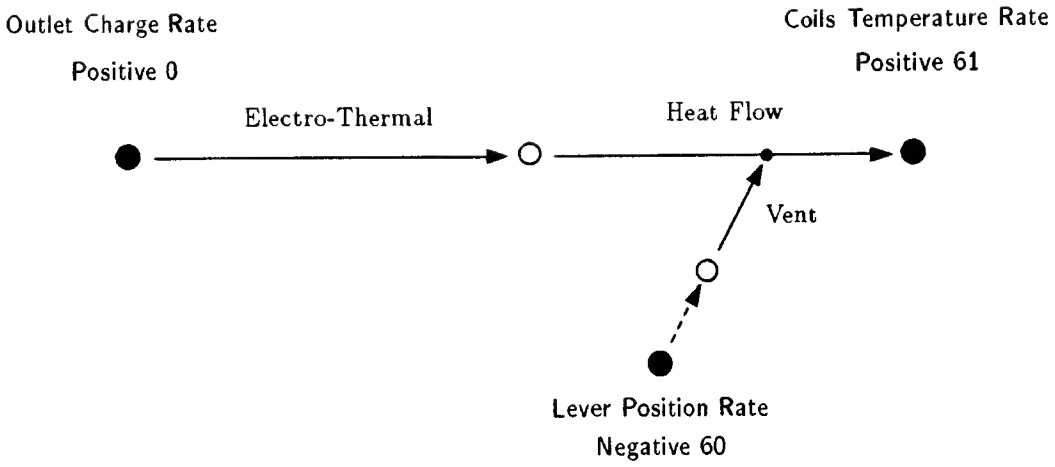


Figure 5.3. Vent hypothesis for coils.

from consideration by reasoning about the magnitude of the gravitational force and the delay associated with temporal integration episodes. The problem is that gravity acts too quickly to account for the observed delay.

Upper bounds on the delays associated with temporal integration episodes are inferred from rates of change and the minimum and maximum possible values of the quantities involved. The rate of change is the propagated value of the magnitude constraint. Limit values for unobservable quantities are taken from defaults specified in mechanism descriptions. After the effects of acceleration are approximated (see Section 6.4.4), the expected time required for a gravity-induced motion to cross the span of a *Switch* or *Vent* is much less than the observed delay. Hypotheses involving gravity still fail to satisfy the delay constraint; on this basis they are removed from consideration.

Reasoning about enablements also leads to hypotheses to explain the upward motion of the lever. The lever is observed to move upwards after having moved downwards. The linear mechanism path hypothesis {·MECHANISM PATH· *Spring-Loading Integration Spring*} nearly accounts for these events: the downward motion of the lever is conjectured to generate a restoring force in a spring; this force increases as the spring is displaced; finally, the force results in motion of the lever in the opposite direction. The problem is that the displacement of the spring and the generation of the restoring force are expected to occur quickly, more quickly than the observed delay between the

downward and upward motions of the lever.

The enablement heuristic is triggered once again; the only constraint violated is the delay constraint. The causal modelling system now attempts to construct enabling mechanism paths which interact with the linear path involving the spring. Again, only those enabling paths which can act slowly enough to account for the observed delay are admitted as hypotheses. One of the admitted hypotheses appears in Figure 5.4.

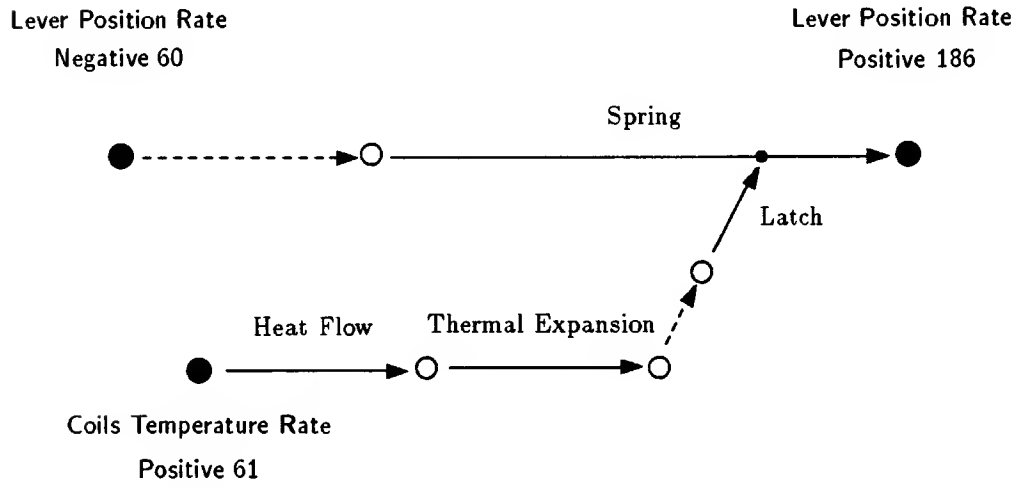


Figure 5.4. Spring and thermally expanding latch hypothesis.

In this hypothesis, the heating of the coils is conjectured to result in motion of a latch through thermal expansion. Eventually, the latch is displaced enough to release the spring, which then moves the lever upwards. The magnitude of the motion associated with thermal expansion is exceedingly small. The time required to move at this rate through the default maximum range of motion for a *Latch* is consistent with the observed delay.

Another hypothesis which accounts for the observed delay between the two motions of the lever appears in Figure 5.5. In this hypothesis, electricity is transformed into motion. This motion displaces a latch until the enabling position is reached and the spring is released.

The range of efficiency associated with the *Electro-Mechanical* mechanism is intended to capture the wide range of stepped-down or stepped-up motors which all operate from electrical power sources of a single magnitude. The

program JACK is able to hypothesize a motorized latch which takes as long as the observed delay between the downward and upward motions of the lever to release a spring.

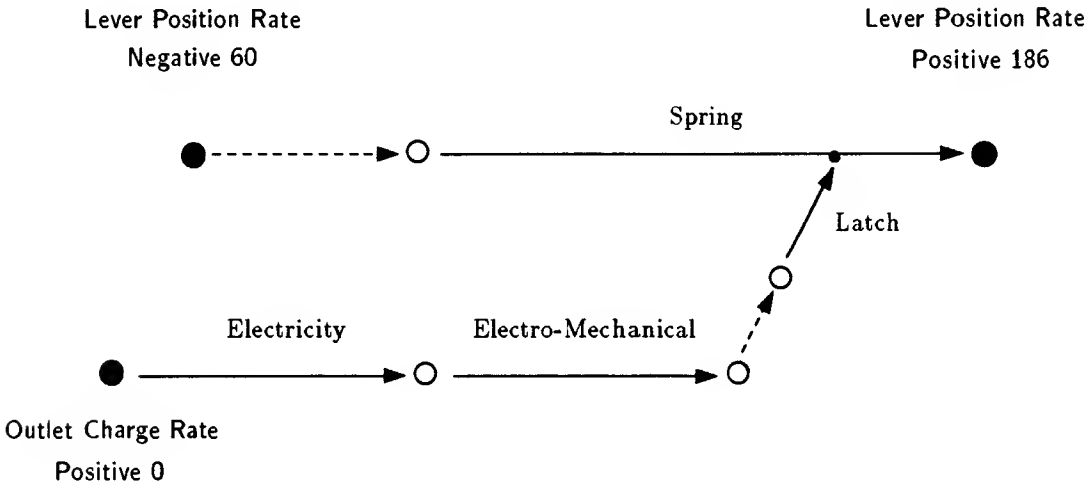


Figure 5.5. Spring and motorized latch hypothesis.

Many enablement interaction hypotheses are pruned because they are unable to explain the observed delay between the opposite motions of the lever. These abandoned hypotheses include latch motions which are gravity-induced and latch motions which are mechanically coupled to the motions of the lever or carriage. All of these hypotheses are inadmissible because the time required to span the maximum range of positions for a latch at the propagated rate of motion is still less than the observed delay.

The causal modelling system admits and prunes the same hypotheses for the opposite motions of the carriage as it does for the opposite motions of the lever. The magnitudes, directions, and times of occurrence of the lever and carriage motions are indistinguishable, and without further information, the program JACK has no basis for conjecturing that one but not the other of these physical objects is spring-loaded and latched.

Most toasters exhibit an annoying inability to reset quickly. If a second piece of toast is placed before the toaster has cooled off appreciably, the lever and carriage will pop up too quickly and the toast will be too light. The

reason for this misbehavior is that the latch on the spring is still partially expanded due to the latent heat in the toaster and will take less time to fully expand to the position at which the spring is released.

The program JACK operates equally well from observations which describe nominal behavior of devices as from observations which describe misbehavior. The approach to modelling implemented in the causal modelling system does not make use of teleological information concerning the intended behavior of a device. Nevertheless, observations of misbehavior can provide opportunities for refining hypotheses. Some proposed device models may provide no means to account for additional observed behavior, whether or not that behavior represents a failure to achieve the intended function of the device.

Figure 5.6 shows a second observation of a toaster. The differences between this observation and the original observation of Figure 5.1 is that the initial temperature of the coils is higher, the delays between the downward and upward motions of the lever and carriage are shorter, and the final darkness of the bread is lighter. The program JACK uses this observation of misbehavior to prune the spring and motorized latch hypothesis of Figure 5.5 while retaining the spring and thermally expanding latch hypothesis of Figure 5.4.

The thermally expanding latch hypothesis is retained because a shorter delay between the downward and upward motions of the lever is both expected and observed. The reasoning which leads to a shorter expected delay is subtle. The delay is attributed to the time needed to move the latch from an initial position to a final position at which enablement occurs. The range of motion of the latch in the two observations is inferred as follows:

The rate of motion of the latch is proportional to the rate of temperature change of the coils; this proportionality is implicit in the assertion of a mechanism path between the two quantities. In the first observation, the temperature of the coils is observed to change from the minimum value *Cold* to the maximum value *Hot*. In the second observation, the initial value of the temperature of the coils is the intermediate value *Warm*. The change of value in the latch position is inferred to be less in the second observation in proportion to the smaller temperature change in the coils. Since the magnitude of the motion is the same in both cases, the delay associated with the displacement of the latch in the second observation is shorter.

The motorized latch hypothesis is pruned because a shorter delay between the downward and upward motions of the lever is not expected, an inference at variance with the shorter observed delay. In this hypothesis, the current at the outlet is the quantity associated with the initial event on the enabling path. The conjectured motion of the latch is proportional to this current. There is no difference in the current in the two observations; hence there is no

		5:00	5:01	5:06	6:06	6:07	10:30
Lever Position Amount			Down			Up	
Lever Position Rate		Negative	Zero		Positive	Zero	
Dial Angle Amount							
Dial Angle Rate							
Carriage Position Amount			Down			Up	
Carriage Position Rate		Negative	Zero		Positive	Zero	
Coils Temperature Amount		Warm			Hot		Off
Coils Temperature Rate			Positive		Zero	Negative	Zero
Bread Appearance Amount		Untoasted			Light		
Bread Appearance Rate		Zero		Positive	Zero		
Outlet Charge Amount							
Outlet Charge Rate							
Earth Gravity Amount							
Earth Gravity Rate							

Figure 5.6. Timeline of second toaster observation.

reason to infer that the latch begins and ends its motion at different positions. The delay for both observations is expected to be the same.

5.1.3 Abstractions and Shortcomings in the Toaster Models

The *Thermal-Expansion* and *Latch* mechanisms appearing in the toaster models constructed by the program JACK are a considerable simplification of how the spring in a toaster actually is released. The force induced in a bimetallic strip by thermal expansion is of too small a magnitude to serve as a robust

release mechanism for a spring. One way in which this force is amplified is by having the expanding bimetallic strip release the support of a small weight in a vertical track. The weight, which moves upward with the carriage, catches on its support when the carriage moves downward. The unsupported weight slides downward under the influence of gravity and displaces a latch on the spring. The falling weight can directly release the spring reliably whereas the feeble bimetallic strip cannot.

A shortcoming shared by all of the proposed causal models of the toaster is the absence of conjectures concerning how the position of the darkness dial contributes to the delay between the downward and upward motions of the lever and carriage. Two factors determine the initial position of the latch in a real toaster: the ambient temperature and the position of the darkness dial, which is mechanically coupled to the latch. The latch must move from this initial position to the enabling position at which the spring is released. The rate of thermal expansion is always the same, hence the initial position of the latch, determined in part by the position of the darkness dial, in turn determines the delay between the opposite motions of the lever and carriage.

The reason for the lack of conjectures about the role of the darkness dial position is straightforward: currently the hypothesis ordering does not include tradeoff interactions—additive interactions where the net magnitude of the effect is non-zero. The causal modelling system can generate additive interaction hypotheses for zero effects; these are the equilibrium hypotheses. The obvious triggering heuristic for tradeoff interactions is one based on the magnitude constraint: suspect an additive interaction when the propagated magnitude is inconsistent with the observed non-zero magnitude. The problem is that the order of magnitude ranges propagated for the magnitude constraint are often so wide that this constraint is not violated when it should be. The alternative of generating additive interaction hypotheses whether or not the magnitude constraint is violated is not in keeping with the principle of Occam's Razor and leads to an explosion of hypothesizing.

I discuss a possible triggering heuristic for tradeoff interactions in Section 6.4.3. This heuristic along with multiple observations of the toaster in which the delay between the downward and upward motions of the lever varies with the position of the darkness dial might enable the program JACK to generate hypotheses concerning the role of the darkness dial position.

5.2 The Pocket Tire Gauge

In this example, I test the program JACK on the surprisingly puzzling pocket tire gauge. The observation for this device appears in Figure 5.7. Initially, the

slide is all the way inside the cylinder and the amount of gas within the tire is stable. Once again, gravity is declared as an available device input. Next, the cylinder of the tire gauge is joined to the tire. Immediately, the amount of gas within the tire decreases and the slide moves out of the cylinder. A short time later, the flow of gas ceases. At the same time, the motion of the slide terminates, an event declared to be a device output. The slide is not at the limit of its range of motion. Some time later still, the cylinder is removed from the tire. The slide does not move again until it is pushed back into the cylinder, an event declared to be another device input.

	0:00	1:00	1:00.1	1:00.2	2:00	2:00.1
Slide Position Amount	G0			G28		G0
Slide Position Rate	Zero		Positive	Zero	Negative	Zero
Tire Amount-of-Gas Amount	P28			P28		
Tire Amount-of-Gas Rate	Zero	Negative		Zero		
Earth Gravity Amount	G					
Earth Gravity Rate	Zero					

Figure 5.7. Timeline of tire gauge observation.

5.2.1 Distinguishing Properties of the Tire Gauge Example

The tire gauge example exemplifies how the design of a device with a relatively simple behavior can be quite puzzling. The tire gauge example turns out to be more complex than the toaster and serves as a more rigorous exercise for the pruning power of the constraints on type, behavior, and structure. The program JACK must conjecture mechanism paths of length four before the target hypothesis emerges. In addition, the tire gauge contains a kind of interaction not found in the toaster: the equilibrium between forces due to air pressure and a spring.

5.2.2 Reasoning About the Tire Gauge

One of the tasks set for the program JACK in the tire gauge example is to explain why the slide stops moving before reaching its limit position. The hypothesis which corresponds to the way a real tire gauge works appears in Figure 5.8. Here the causal modelling system conjectures that an equilibrium has been achieved. The two opposing contributions which make up the equilibrium are a pneumatically-induced motion of a hidden physical object due to the flow of gas from the tire, and a spring-induced motion of the same physical object in the opposite direction due to displacement of a spring by the moving object, resulting in a restoring force in the opposite direction to the displacement.

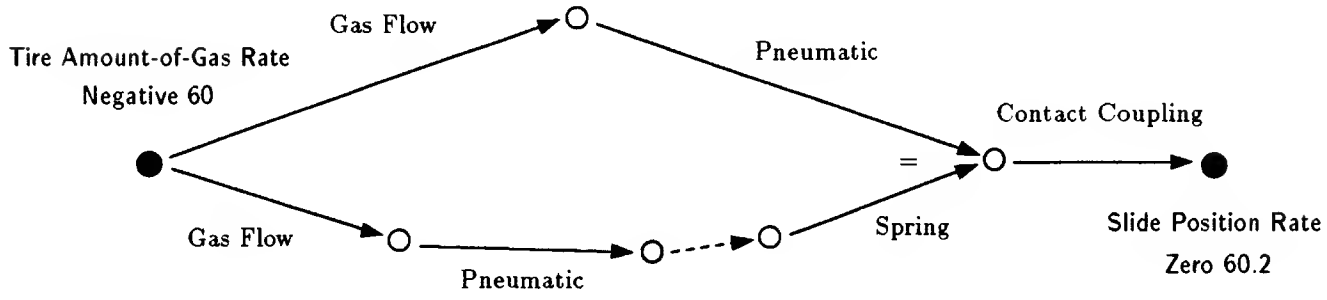


Figure 5.8. Spring hypothesis.

The triggering heuristics for suspecting equilibrium and disablement situations are the same. Not surprisingly, the program JACK is able to generate disablement hypotheses to explain the halting of the motion of the slide. One of these hypotheses appears in Figure 5.9. This proposed causal model for the tire gauge also involves pneumatically-induced motion of a hidden physical object. However, in this case the motion of the hidden object displaces not a spring but a valve. When the valve is closed, the flow of gas is disabled, and the motion of the slide—transmitted along a mechanical coupling from the hidden object—also stops. Thus an impulse of displaced gas is responsible for the start-and-stop motion of the slide.

An alternate disablement hypothesis generated by the causal modelling system proposes that the pneumatic motion of the hidden object, rather than closing a valve which disables the flow of gas, instead engages a latch which directly arrests the motion of the slide.

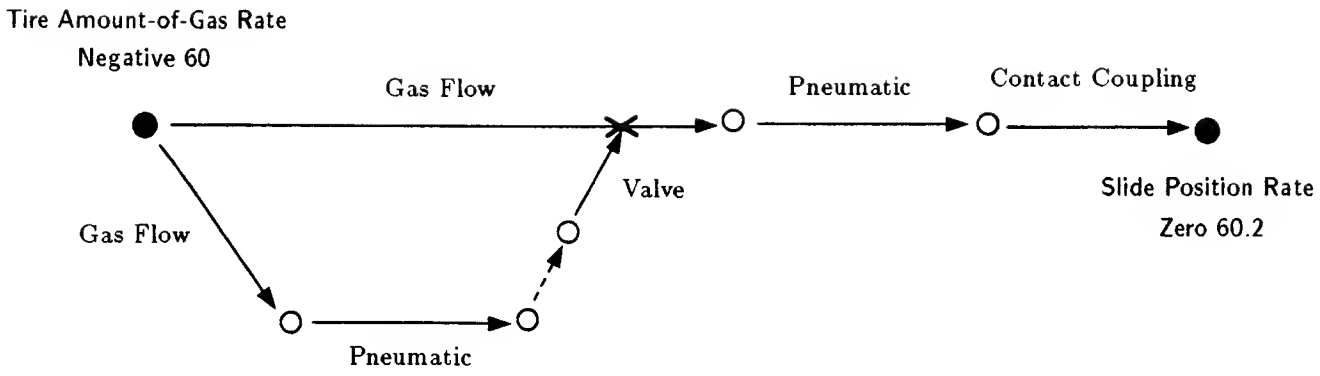


Figure 5.9. Impulse hypothesis.

Unfortunately, additional observations of the tire gauge describing the motion of the slide for different input tire pressures do not serve to prune these "impulse" hypotheses. Let us examine the reasoning involved closely, using the case of a second observation involving a greater amount of gas inside the tire. For both the spring hypothesis and the impulse hypothesis, a greater rate of gas flow from the tire is propagated via the magnitude constraint to a greater rate of motion for the hidden object. The time required for the displacement of the spring may be shorter, the same, or longer: the rate of motion is greater but a greater displacement of the spring is required to achieve equilibrium. On the other hand, the time required for the displacement of the valve in the impulse hypothesis is strictly shorter: the valve becomes closed at the same position and the rate of motion is greater. Either inference about the change in delay until equilibrium or disablement is achieved is compatible with the observed shorter duration of the motion of the slide. Furthermore, this shorter duration combined with the greater rate of motion of the slide is consistent with the greater displacement of the slide.

Another opportunity to reason about the spring and impulse hypotheses is afforded by the part of the tire gauge observation which describes how the slide, which had been stationary, continues to be motionless when the cylinder of the tire gauge is removed from the tire. This part of the observation, while not distinguishing the two hypotheses, does shed some light on the nature of the mechanical coupling between the hidden object and the slide conjectured in both hypotheses.

The false {·RELATION· *Tire Joined-To Cylinder*} relation violates the medium constraint for the *Gas-Flow* mechanism in both hypotheses. For the impulse hypothesis, this results in both the primary path and the disabling path in a disablement interaction becoming inactive. The prediction of no expected effect (see Table 4.3) is consistent with the observation of the slide remaining motionless.

The reasoning for the spring hypothesis is considerably more subtle. Both halves of the proposed equilibrium become inactive because the now-unsupported *Gas-Flow* mechanism appears as the first mechanism along both interacting paths. However, and this is a key point, the two mechanism paths do not become inactive at the same time. The delay along the mechanism path which contains the spring is longer. Just as time is required to displace the spring and achieve the equilibrium state, so time is required to unload the spring and remove this influence on the position of the hidden object. There is an interval during which the pneumatic half of the equilibrium interaction has become inactive while the spring half is still active. The program JACK is able to infer this broken equilibrium from the unequal delays along the two mechanism paths and predicts, according to Table 4.4, that the hidden object moves in the direction opposite to its original motion.

The task now is to explain how the slide need not move despite the conjectured motion of the hidden object inside the tire gauge. Three types of mechanical couplings between the hidden object and the slide are proposed by the causal modelling system as part of the spring and impulse hypotheses: the *Rigid-Coupling*, *Contact-Coupling*, and *Forward-Ratchet* mechanisms. The *Rigid-Coupling* is predicted to be active and is inconsistent with the motionless slide. The slide should move into the cylinder along with the hidden object. The *Contact-Coupling* mechanism is predicted to be inactive because the alignment constraint is violated: the position of the hidden object is greater than, not less than, the position of the slide along the direction of motion. In other words, the hidden object cannot pull the slide. This mechanism can explain the stationary slide. The *Forward-Ratchet* mechanism also is predicted to be inactive because the bias constraint is violated: the motion is not in the positive direction—the only direction allowed. This mechanism also is compatible with the slide remaining at rest.

The *Forward-Ratchet* mechanism ultimately is eliminated when the slide is pushed back manually into the cylinder. This observation is inconsistent with the prediction that the slide will not move in this direction.

5.2.3 Abstractions and Shortcomings in the Tire Gauge Models

The abstracted account of classical mechanics inherent in the representation of mechanical coupling mechanisms is the source of one of the shortcomings in the causal models proposed by the program JACK to explain the events in the observation of the tire gauge. This account is more Aristotelian than Newtonian. In particular, there is no notion of momentum. Thus a sufficient explanation for the halting of the slide's motion is the equilibrium or disablement which stops the motion of the hidden object coupled to the slide only through contact. There is no apparent deficiency in this explanation which might be removed in a more complete model which includes friction.

The causal modelling system explains events involving quantities moving to their zero values only through equilibrium interactions or disablement interactions. There is a third possibility: an intermediate, hidden quantity may reach a limiting value, with this forced cessation of change propagating to the observable effect event. A limit value hypothesis can explain the halting of the motion of the slide in the tire gauge observation. The flow of gas from the tire may result in pneumatic motion of a hidden physical object. However, this object may be constrained to move only so far. The slide, mechanically coupled to this object, stops moving when the object stops moving. This "stop" hypothesis is shown in Figure 5.10.

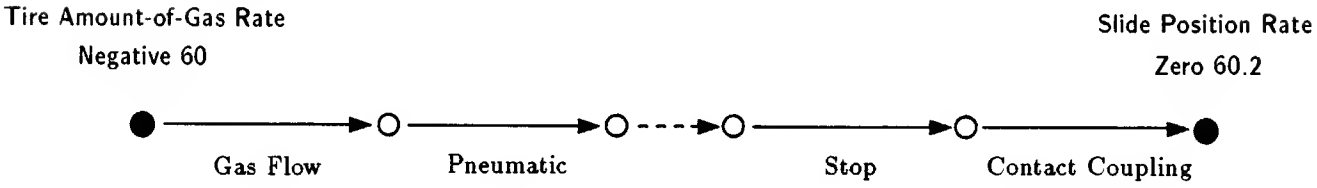


Figure 5.10. Stop hypothesis.

The hidden object moves to a limit position in the depicted temporal integration episode. The *Stop* mechanism records that the rate of motion is zero after this episode.

However, the stop hypothesis cannot explain how the slide moves further out of the cylinder when the flow of gas from the tire is greater. The prediction for any gas flow rate is that the hidden object moves to its unchanging limit value and stops. The slide also moves to the same position each time. The

final position of the slide cannot be proportional to the input gas flow rate, as it is observed to be. This hypothesis would be pruned on the evidence of additional observations.

Another hypothesis which is beyond the capability of the current version of the program JACK combines the limit value form of hypothesis with the concept of momentum. This hypothesis is an extension of the stop hypothesis. When the hidden physical object reaches its limit value, it stops moving. However, the slide, not being attached to this object, not having reached the end of its range of motion, and operating under the physical principle of momentum, continues to move out of the cylinder. Friction ultimately terminates this motion. This "throw" hypothesis is shown in Figure 5.11.

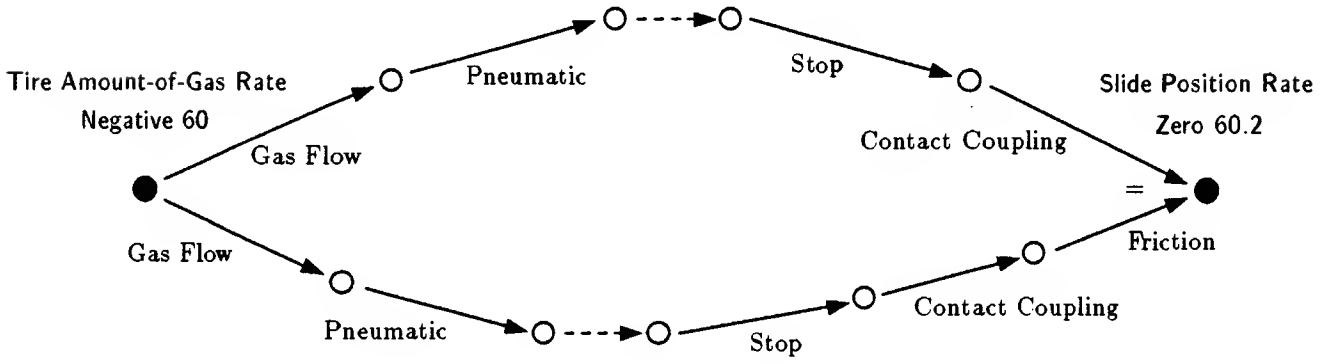


Figure 5.11. Throw hypothesis.

Continued motion due to momentum can be captured by propagating the event node following the *Pneumatic* mechanism past the *Stop* mechanism without change. The *Friction* mechanism depicted is similar to the *Spring* mechanism; in both motion engenders a restoring force in the opposite direction. However, in the case of friction, the magnitude of the force is proportional to the rate of motion, not to the displacement.

The mechanical couplings which can be incorporated into the throw hypothesis are those which do not involve attachment. Continued motion due to momentum is suppressed by attachment; the slide would stop moving when the hidden object stopped moving. Both the *Rigid-Coupling* mechanism and the *Forward-Ratchet* mechanism—which is simply a rigid coupling with a bias—involve attachment and would be inadmissible. The *Contact-Coupling*

mechanism does not and would permit the slide to continue to move outward via momentum.

The throw hypothesis would not be pruned by additional observations of the tire gauge involving different rates of gas flow from the tire. A greater input gas flow rate results in a greater velocity for both the hidden object and the slide. Due to this greater velocity, the slide moves further before being halted by friction. The throw hypothesis can explain how the final position of the slide is proportional to the input gas flow rate.

5.3 The Bicycle Drive

In this example, I test the program JACK on a simplified version of the old style of bicycle drive distinguished by the coaster brake. The coaster brake is engaged by pedaling backward. The observation of a bicycle drive appears in Figure 5.12. Initially, the pedal, sprocket, and hub of the back wheel are all stationary. Then the pedal begins to rotate forward; this event is declared to be a device input. At the same time, the sprocket begins to rotate in the same direction. Finally, after a slight delay, the hub also rotates forward in an event declared to be a device output. Later, the pedal stops rotating; this event is another device input. Simultaneously, the sprocket stops rotating but the hub continues to rotate. Later still, the pedal is rotated in the opposite direction—yet another input. The sprocket instantly rotates in the same direction. Then the pedal and sprocket stop rotating at the same time. The hub also stops rotating at this time. This last event is an output of the device.

5.3.1 Distinguishing Properties of the Bicycle Drive Example

The type constraint is partially compromised in the bicycle drive example because all of the quantities in the observation are angle quantities. This example helps to reveal the pruning power of the remaining constraints on behavior and structure. The linkages in the bicycle drive which engage the back wheel and the brake are one-way linkages which operate in opposite directions. The alignment and bias constraints, which support reasoning about one-way behavior, are one of the keys to the performance of the program JACK on this example. Furthermore, the one-way nature and independence of the drive and brake linkages can be inferred only by comparing multiple instances of events. This example serves also to test the hypothesis refinement procedure.

	0:00	1:00	1:01	1:10	1:20	1:21
Pedal Angle Amount	Top			Front		Top
Pedal Angle Rate	Zero	Positive		Zero	Negative	Zero
Sprocket Angle Amount	Front			Bottom		Front
Sprocket Angle Rate	Zero	Positive		Zero	Negative	Zero
Hub Angle Amount	Back					Bottom
Hub Angle Rate	Zero		Positive			Zero

Figure 5.12. Timeline of bicycle drive observation.

5.3.2 Reasoning About the Bicycle Drive

The bicycle drive example involves two modelling tasks: (1) hypothesizing how the angle of the back wheel hub is made to increase in the forward direction, and (2) hypothesizing how this rotary motion is made to cease. The causal modelling system has a repertoire of five rotary linkages to consider. There is an attachment coupling: *Rigid-Rotary-Coupling*. There is a push-but-not-pull coupling: *Contact-Rotary-Coupling*. There is a pull-but-not-push coupling: *Non-Rigid-Rotary-Coupling*. There are two ratchet couplings, one in each direction: *Forward-Rotary-Ratchet* and *Backward-Rotary-Ratchet*. Finally, there is a mechanism for enabling and disabling rotary motion: *Rotary-Latch*.

Two of the rotary linkages fail to support hypotheses in which the forward rotation of the sprocket is the cause of the forward rotation of the hub. The *Rigid-Rotary-Coupling* mechanism fails to explain the observed delay and the *Backward-Rotary-Ratchet* mechanism can explain only backward rotations.

The program JACK refines the set of remaining hypotheses against other events involving the angle of the sprocket. One of these events describes a backward rotation of the sprocket. The causal modelling system predicts the outcome of this event for each hypothesis according to Table 4.1 and verifies these predictions against the actually observed events in the timeline.

The *Forward-Rotary-Ratchet* hypothesis is dismissed because the bias

constraint is violated; rotation in the backward direction is not possible. The *Contact-Rotary-Coupling* and *Non-Rigid-Rotary-Coupling* mechanisms are both predicted to be inactive. In both cases, the alignment constraint plays a central role. In the original *Contact-Rotary-Coupling* hypothesis the position of the sprocket was asserted to be less than the position of the hub along the direction of motion. The direction of motion is now reversed and the sprocket, which had pushed the hub forward, cannot now pull it backward. The position of the sprocket was asserted to be greater than the position of the hub in the original *Non-Rigid-Rotary-Coupling* hypothesis. The sprocket, which had pulled the hub forward, cannot now push it backward. Both of these hypotheses are retained because no backward rotation of the hub is expected, and none is observed. The models which are consistent with the forward and backward rotations of the sprocket are summarized in Figure 5.13.

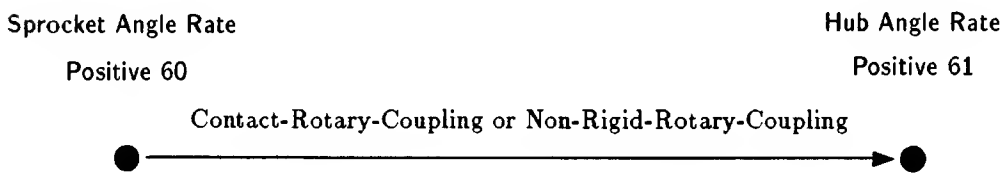


Figure 5.13. One-way drive linkages.

The event describing the halted rotation of the hub triggers the heuristics for equilibrium and disablement situations. An interacting mechanism path is suspected either to balance or inhibit the rotation of the hub caused by the rotation of the sprocket. No equilibrium hypotheses are generated but a number of disablement hypotheses satisfy the constraints. The disabling mechanism path in all of these hypotheses begins at the backward rotation of the pedal and consists of a rotary linkage and the *Rotary-Latch* mechanism. In each disablement hypothesis, the *Rotary-Latch* mechanism interacts with the rotary linkage between the sprocket and the hub. The only rotary linkage which is not proposed for the disabling path is the *Forward-Rotary-Ratchet* mechanism. This mechanism participates only in forward rotations.

The program JACK refines this set of disablement hypotheses using reasoning similar to that used to refine the set of hypotheses which explain the forward rotation of the hub. Predicted outcomes of the event describing forward rotation of the pedal are generated for each of the proposed disablement

interactions using Table 4.3. These predictions are verified against the actual observation of the bicycle drive. The prediction for the disablement path involving a *Rigid-Rotary-Coupling* mechanism is that the disablement path is active and the hub stops rotating. This prediction is unverified and this disablement hypothesis is pruned. The hypothesis involving a *Backward-Rotary-Ratchet* mechanism is dismissed because the bias constraint is violated; rotation in the forward direction is not possible.

The other two proposed disablement paths are predicted to be inactive because of the alignment constraint. In the hypothesis involving a *Contact-Rotary-Coupling* mechanism, the pedal is conjectured to be pushing a hidden physical object which it cannot now pull. Similarly, in the hypothesis involving a *Non-Rigid-Rotary-Coupling* mechanism, the pedal is conjectured to be pulling a hidden object which it cannot now push. Each of these disablement hypotheses is retained because no inhibiting of the hub rotation is expected from the forward rotation of the pedal and none is observed. See Figure 5.14.

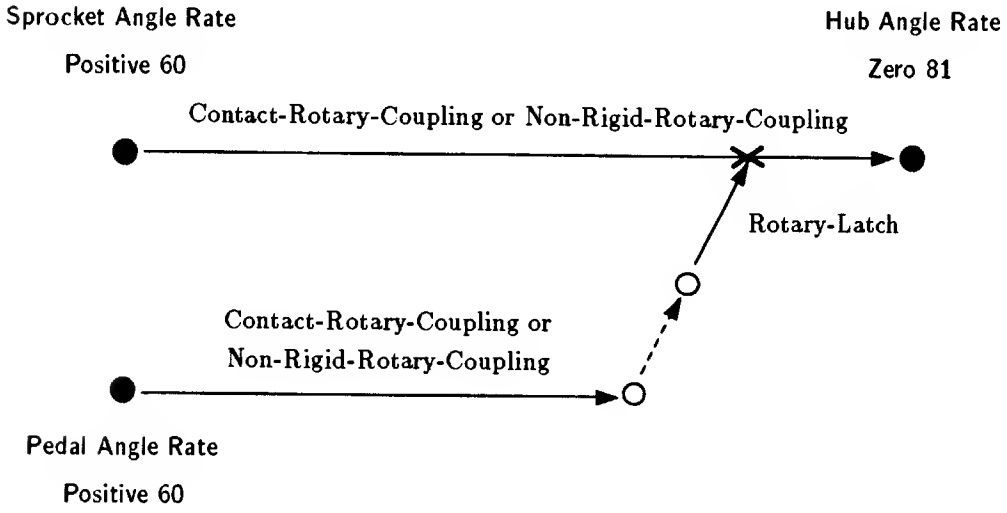


Figure 5.14. One-way drive and brake linkages.

5.3.3 Abstractions and Shortcomings in the Bicycle Drive Models

The reasoning employed by the program JACK in refining the set of models for the bicycle drive does not incorporate the concept of momentum—a short-

coming exhibited also in the modelling of the tire gauge. Because of this limitation, a *Rigid-Rotary-Coupling* hypothesis for the drive linkage between the sprocket and the hub cannot be dismissed on the basis of the continuing forward rotation of the hub after the forward rotation of the sprocket stops. Rather, this hypothesis is pruned because it cannot account for the observed delay between the two rotations. This reasoning, while successful in this case, would be buttressed by knowledge of the concept of momentum.

The *Rotary-Latch* mechanism used by the causal modelling system to explain how the hub of the back wheel stops rotating abstracts considerably away from how a real bicycle coaster brake system works. The *Rotary-Latch* mechanism simply maps the amount of one angle quantity to the rate of another angle quantity. The explanation in the bicycle drive example is that the backward rotation of the pedal moves the rotary latch to its zero value, resulting in a zero value for the rate of the hub angle.

A real coaster brake works in the following way: the backward rotation of the pedal is transmitted to a long strip of material through a one-way linkage. This strip of material is coiled against the rim of the hub of the back wheel. When pulled, it wraps more closely against the hub rim. Friction between the strip and the hub slows the rotation of the hub and can halt it altogether.

The structural and geometrical information needed to conjecture a better approximation of a coaster brake system is not available—the same limitation seen in the modelling of a latch in the toaster. In particular, the representation of the *Rotary-Latch* mechanism does not support reasoning about the inward coiling behavior of a flat physical object which is anchored at one end and pulled from the other. Nor does it support reasoning about the change in linear distance between the inner surface of the coiling physical object and the outer surface of the object to which it is attached—a change which results in contact. The program JACK is unable to generate explanations which involve complex structural and geometrical constraints.

5.4 The Refrigerator

In this example, I test the program JACK on an idealized version of a refrigerator. The observation of a refrigerator is shown in Figure 5.15. Initially, the temperature of the interior of the refrigerator is *Cold* and the temperature of the exterior is *Ambient*. Current at the outlet is declared to be an available input. Shortly thereafter, the temperature of the interior rises. Some time later, the temperature inside the refrigerator begins to decrease; this is declared to be an output of the device. Shortly after that, the temperature outside the refrigerator begins to increase; another output.

	0:00	0:01	1:00	1:01
Interior Temperature Amount	Cold			
Interior Temperature Rate	Zero	Positive	Negative	
Exterior Temperature Amount	Ambient			
Exterior Temperature Rate	Zero			Positive
Outlet Charge Amount	On			
Outlet Charge Rate	Positive			

Figure 5.15. Timeline of refrigerator observation.

5.4.1 Distinguishing Properties of the Refrigerator Example

The refrigerator example is the most complex device example on which the program JACK has been tested. Mechanism paths of length four are needed to generate the target hypotheses. Furthermore, this example tests the ability of the causal modelling system to reason about hidden inputs and cycles. The hidden inputs inside a refrigerator are heat gains and heat losses in a refrigerant which acts as a heat carrier. This refrigerant circulates in a cycle in which the amount of heat alternately gained and lost remains bounded.

5.4.2 Reasoning About the Refrigerator

In one of the causal models generated by the program JACK to explain the observation of a refrigerator, the two halves of the cycle of operations within a compression refrigerator are identified, albeit approximately. This model is depicted in Figure 5.16. One mechanism path explains how the interior of the refrigerator gets colder: Current from the outlet is transformed into motion via the *Electro-Mechanical* mechanism. This motion results in a pressure decrease through the *Expansion* mechanism. This pressure decrease, once a threshold is crossed, generates a heat loss at the refrigerator interior via the *Evaporation* mechanism.

A different mechanism path explains how the exterior of the refrigerator gets warmer. Along this path, electrically-induced motion results instead in a pressure increase through the *Compression* mechanism. This pressure increase, after a threshold is crossed, results in a heat gain at the refrigerator exterior via the *Condensation* mechanism.

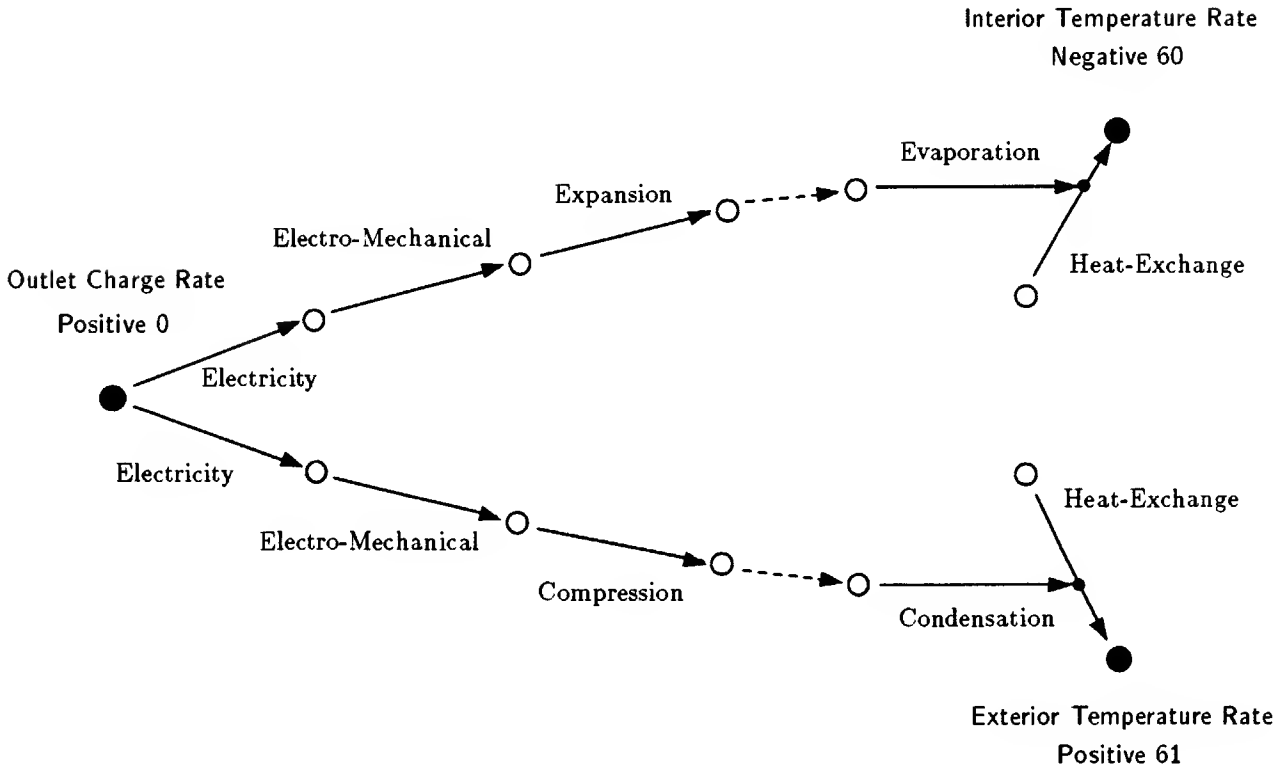


Figure 5.16. Compression refrigerator hypothesis.

Both of these mechanism paths involve hidden inputs. The *Evaporation* and *Condensation* mechanisms along the paths provide incomplete explanations; they describe how pressure changes can result in a heat loss or heat gain but do not indicate where the heat goes to or where it comes from. The *Evaporation* and *Condensation* mechanisms can only appear in enablement and disablement interactions. Without the missing contribution representing the heat sink or heat source, they do not provide explanations for heat losses or heat gains.

Whenever the program JACK constructs a linear mechanism path which contains one of these "interaction-only" mechanisms, the heuristic for suspecting hidden input situations is triggered. A hidden input hypothesis is constructed by propagating forward from the cause event in the normal manner to the required point of interaction, propagating backward from the effect event to the same point, inverting the enablement interaction at this point, and extending the hidden input path backward through one mechanism. In the refrigerator hypothesis shown in Figure 5.16, this procedure results in conjecturing *Heat-Exchange* mechanisms interacting with the *Evaporation* and *Condensation* mechanisms, and inferring event nodes describing a hidden heat sink and a hidden heat source. These hypotheses must be admitted as long as any values can be propagated to the event nodes describing the hidden inputs; these event nodes cannot, by definition, be compared to observable events.

The model for a refrigerator proposed by the causal modelling system implies an internal heat source and an internal heat sink. Refrigerators plausibly could be designed in this way. There could be two additional inputs besides electricity: a gas line and a fluid line. The gas could be condensed on demand, giving off heat to the exterior and the fluid could be evaporated on demand, taking up heat from the interior. Of course, heating of the exterior is entirely gratuitous in this design. But the proposed model is reasonable, given that the program JACK is not provided with the teleological information that cooling of the interior is the only intended function of a refrigerator and heating of the exterior is merely a side effect.

An alternate model generated by the causal modelling system approximates the absorption type of refrigerator. This model is shown in Figure 5.17. The only difference between this model and the one appearing in Figure 5.16 is the means by which a pressure increase is achieved along the mechanism path which terminates in the *Condensation* mechanism. Instead of being achieved mechanically via the *Compression* mechanism, the pressure increase is brought about, curiously enough, by raising the temperature through the *Thermal-Compression* mechanism.

This explanation is actually accurate, as far as it goes. The pressure increase which forces condensation may be brought about either through the manipulation of volume or through the manipulation of temperature. However, this explanation introduces another heat source of arbitrary capacity, in addition to the heat source. In a real absorption refrigerator, some of this additional heat is transferred to the refrigerant vapor returning from the evaporator before it is directly heated, to assist in forcing condensation. This aspect of the operation of an absorption refrigerator goes unmodelled by the program JACK in the hypothesis of Figure 5.17.

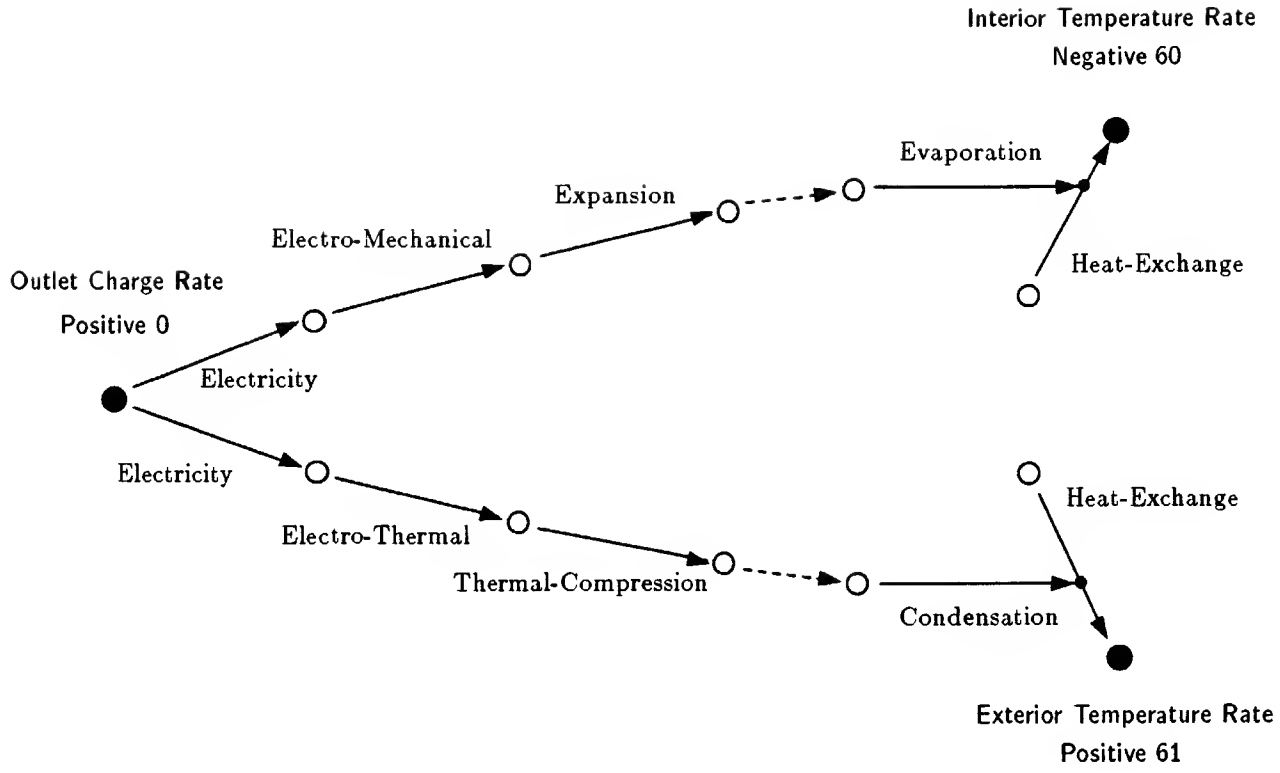


Figure 5.17. Absorption refrigerator hypothesis.

The causal modelling system, not being constrained to generate models where the heating of the exterior is a side effect of the cooling of the interior, also proposes a model for the refrigerator involving direct electrical heating of the exterior. This hypothesis is depicted in Figure 5.18. This model for a refrigerator involves only a single hidden input. A refrigerator designed in this way would need a fluid line to seed the evaporation process which cools the interior but would not require a gas line to seed condensation. The heating of the exterior is traceable to a known input to the device: the current at the outlet. In an ordering based only on the number of hidden inputs, the direct heating hypothesis is preferable to the compression refrigerator hypothesis. However, the compression refrigerator hypothesis can be improved by introducing a cycle, whereas the direct heating hypothesis cannot.

The heuristic for suspecting cycle situations is triggered by conjectured

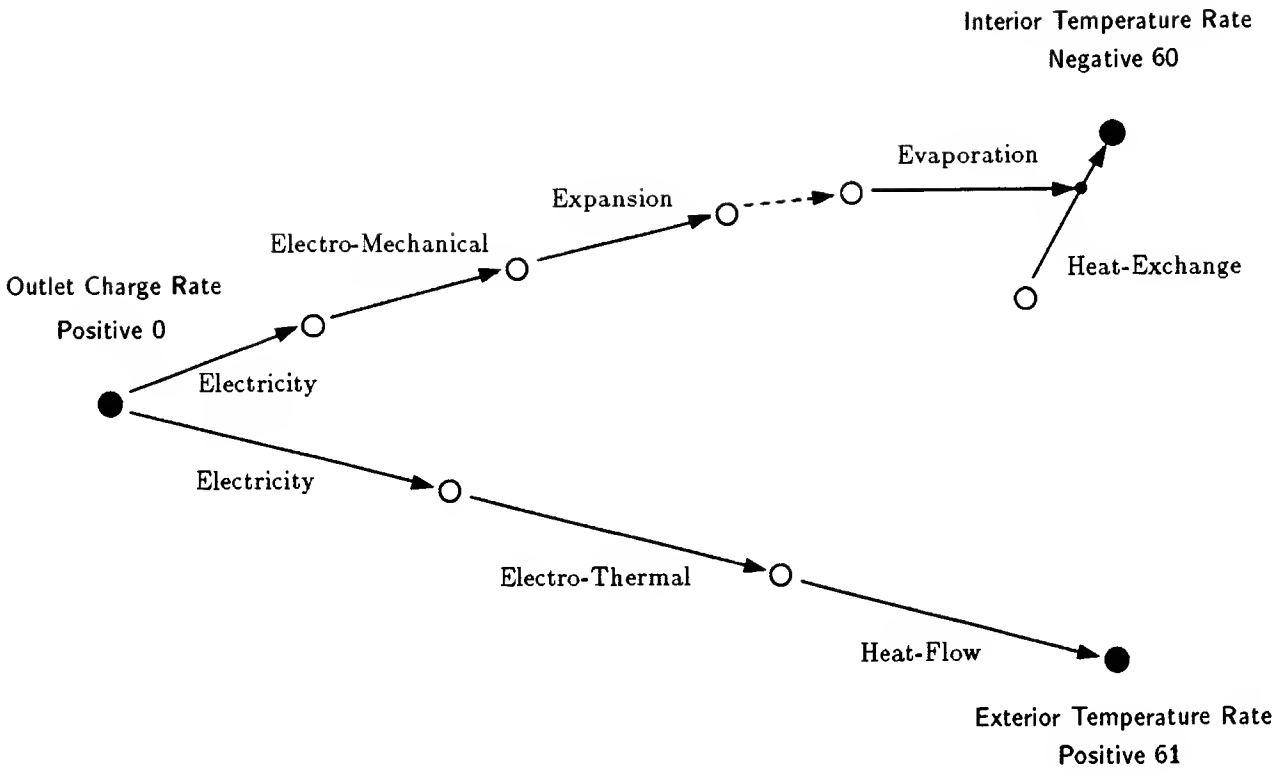


Figure 5.18. Direct heating hypothesis.

hidden inputs of opposite sign. The causal modelling system constructs cycle hypotheses by additively combining pairs of hidden inputs in much the same way that the contributions of separate mechanism paths are combined in proposed equilibrium interactions. The only cycle hypotheses admitted are those for which the net change around the proposed closed loop may be zero and for which the separate hidden inputs involve the same physical object. These cycle hypotheses show how the conjectured sources and sinks may be avoided.

A cycle hypothesis admitted by the program JACK is shown in Figure 5.19. In this hypothesis, the internal heat gain associated with the *Evaporation* mechanism and the internal heat loss associated with the *Condensation* mechanism form two halves of a cycle and do not accumulate. This hypothesis captures in part the synergy in the actual design of compression refrigerators.

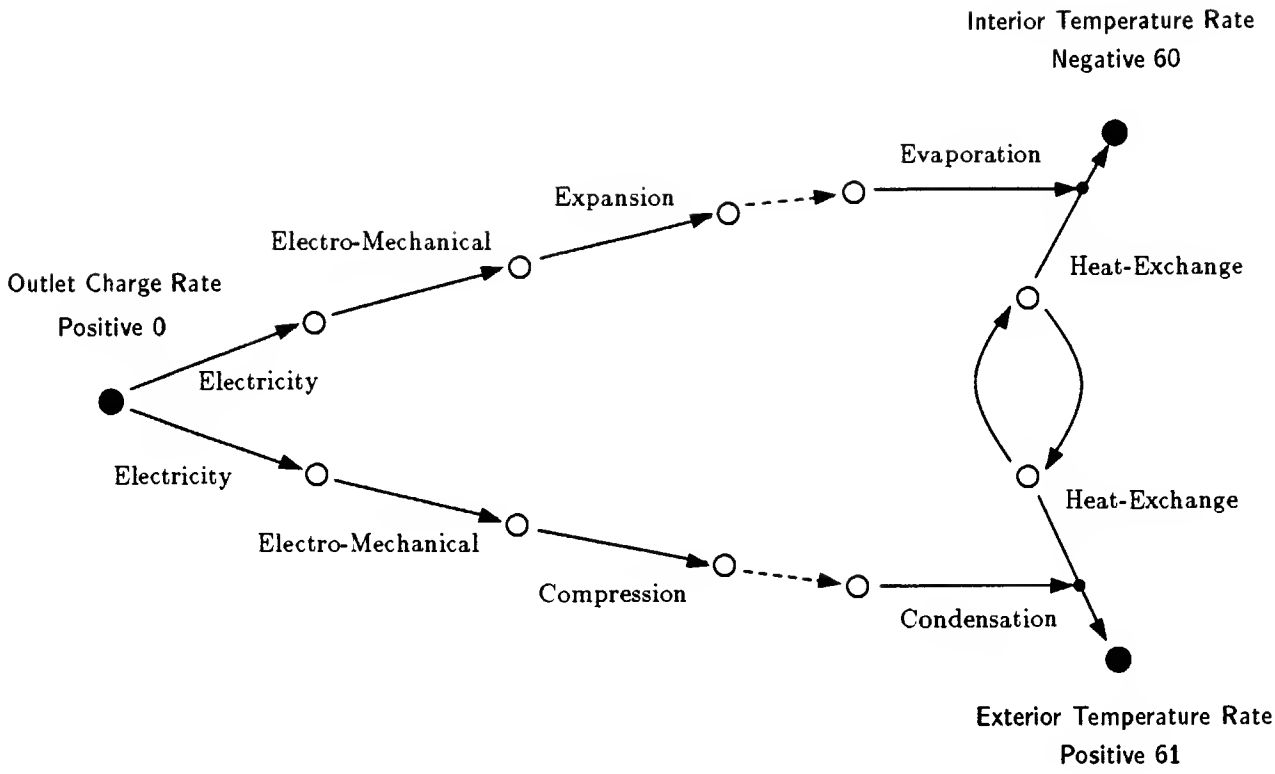


Figure 5.19. Refrigerant cycle hypothesis.

The causal modelling system is unable to form a cycle hypothesis for the direct heating model of the refrigerator. There is no heat sink which can offset the heat source associated with the *Evaporation* mechanism. In an ordering based on the number of hidden sources and sinks, the compression refrigerator hypothesis with a cycle is preferable to the direct heating hypothesis.

5.4.3 Abstractions and Shortcomings in the Refrigerator Models

Several of the mechanisms appearing in the proposed models for the refrigerator represent constraints which are consequences of the Ideal Gas Law: $PV = nRT$. For example, the *Compression* and *Expansion* mechanisms capture the volume-pressure relationship. The *Thermal-Compression* mechanism captures the temperature-pressure relationship. The remaining variables in

the Ideal Gas Law are not mentioned in these mechanism descriptions. Other mechanism descriptions also contain hidden parameters: the expansion coefficient in the *Thermal-Expansion* mechanism, the threshold temperature and pressure in the *Condensation* and *Evaporation* mechanisms.

Two consequences result from hidden parameters: Typically, a range of reasonable values for hidden parameters are encoded into the efficiency slot of a mechanism description. This range, which can be quite wide, compromises the utility of the magnitude constraint. Furthermore, an assumption that hidden parameters are being kept constant may be invalid.

Some of the reasoning employed by the program JACK in the modelling of the refrigerator incorporates the constancy assumption for hidden parameters. Specifically, the temperatures at which phase changes occur in the *Evaporation* and *Condensation* mechanisms are not computed as a function of pressure; rather, enablement of these processes is associated with a fixed default value for pressure. Effectively, evaporation or condensation is inevitable, given a monotonic pressure decrease or increase. This representation of these processes misses some of the subtlety of the principles involved, particularly in the absorption refrigerator where condensation is forced by raising the boiling point by increasing the pressure, even as the temperature is raised. The causal modelling system is not misled by the hidden parameter assumption in this case, but nevertheless the rendering of the physical principles upon which the reasoning is based is incomplete.

Conservation is one of the most powerful of physical principles. This is the main principle behind the construction of cycle hypotheses in the reasoning of the causal modelling system. The conservation principle states that the total amount of any quantity is a physical system does not change. In a closed system, any increases in a quantity in one part of the system are offset by balancing decreases in another part of the system. In an open system, any source of a quantity entering the system is offset by a sink of that quantity leaving the system; the total amount of the quantity does not arbitrarily increase or decrease within the system.

A proposed hidden input is a conjectured source or sink to an open system. A cycle hypothesis demonstrates how the increase or decrease associated with a proposed hidden input can be balanced elsewhere in the system; how a hypothesized source or sink to an open system can be successfully subsumed into a closed system. This conservation reasoning applies to both energy quantities, e.g., *Temperature*, and mass quantities, e.g., *Amount-of-Fluid*.

Cycles within a device can produce a form of synergy which illustrate the conservation principle—gains in one part of a cycle are balanced by losses in another part. There are other cyclic phenomena in physical systems which

are not modelled by the program JACK. In particular, the causal modelling system does not reason about iteration. The cooling of the interior of a refrigerator is the result of many "pulses" of heat loss due to the evaporation of refrigerant. The program JACK does not reason about how small changes can be accumulated into large changes through iteration. The models proposed by the causal modelling system describe only a single cycle of operations within a refrigerator.

5.5 The Home Heating System

In this example, I test the program JACK on a simplified version of a home heating system. The observation of a home heating system appears in Figure 5.20. Initially, the temperatures of a furnace, a radiator and a room are all stable. Current at an outlet and gravity are declared to be available inputs. First the temperature of the room begins to fall. Some time later, the temperature of the furnace begins to increase. Shortly thereafter, this temperature increase ceases. Still later, the temperature of the radiator and the temperature of the room rise simultaneously. The change in room temperature is declared to be an output of the system. Finally, the temperatures of the radiator and the room stop changing.

5.5.1 Distinguishing Properties of the Home Heating Example

The home heating system example is moderately complex compared to the other device examples. As in the refrigerator example, the target hypothesis involves a hidden input—in this case, the water which transports heat from the furnace to the radiator. This hypothesis involves three interacting mechanism paths. Three interacting mechanism paths imply up to two hidden inputs; this example also serves to reveal the loss of pruning power which results when the causal modelling system is given free reign to conjecture hidden inputs. Finally, this example exposes some of the difficulties involved in representing and reasoning about mass quantities such as fluids.

5.5.2 Reasoning About the Home Heating System

The target hypothesis for the home heating system, successfully generated by the program JACK, is shown in Figure 5.21. The mechanism of heat transfer from the furnace to the radiator is simply *Heat-Flow*. However, two enablement interactions support this heat transfer. There is a *Fluid-Heat-Transport* mechanism; in this mechanism a fluid flow supports heat transfer

	0:00	0:01	6:00:00	6:01:00	6:03:00	6:06:00	
Furnace Temperature Amount	Off				On		
Furnace Temperature Rate	Zero			Positive	Zero		
Radiator Temperature Amount	Cold						
Radiator Temperature Rate	Zero					Positive	
Room Temperature Amount	Nice		Cool				
Room Temperature Rate	Zero	Negative					
Outlet Charge Amount	On						
Outlet Charge Rate	Positive						
Earth Gravity Amount	G						
Earth Gravity Rate	Zero						

	6:06:10	6:09:00	6:16:10
Furnace Temperature Amount			
Furnace Temperature Rate			
Radiator Temperature Amount		Hot	
Radiator Temperature Rate		Zero	
Room Temperature Amount			Nice
Room Temperature Rate	Positive		Zero
Outlet Charge Amount			
Outlet Charge Rate			
Earth Gravity Amount			
Earth Gravity Rate			

Figure 5.20. Timeline of home heating system observation.

between two locations. There is also a *Pump* mechanism; in this mechanism motion initiates a fluid flow. The hidden input in this hypothesis is the missing fluid source at the enablement interaction involving the *Pump* mechanism. The causal modelling system proposes a *Fluid-Exchange* mechanism at this point of interaction.

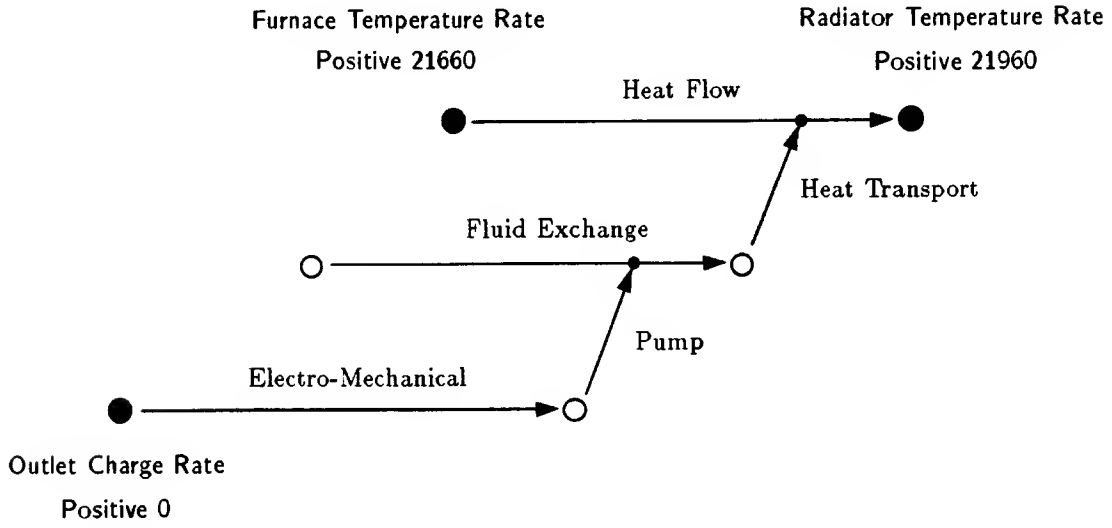


Figure 5.21. Pumped heat transport hypothesis.

Several of the other home heating models proposed by the program JACK are variations of the hypothesis of Figure 5.21. In one of these alternate hypotheses, *Fan* and *Gas-Heat-Transport* mechanisms are substituted for the *Pump* and *Fluid-Heat-Transport* mechanisms, respectively. This hypothesis captures an abstraction of a home heating system based on circulating steam, rather than circulating hot water.

With the freedom to conjecture up to two hidden inputs, the program JACK is able to generate some unusual hypotheses. One of these is shown in Figure 5.22. In this hypothesis, the delay between the heating of the furnace and the heating of the radiator is attributed to the time required to open a vent. The vent is moved pneumatically via a gas flow driven by a fan. The hidden inputs in this hypothesis are the gas source and the origin of the fan's motion.

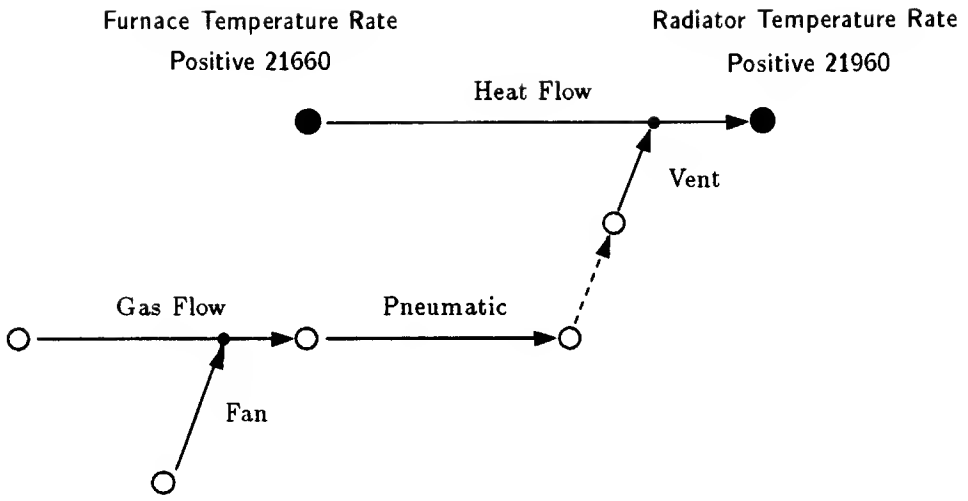


Figure 5.22. Pneumatic vent hypothesis.

The observed delay before the temperature of the furnace begins to rise is on the order of several hours. This observation might describe a home heating system which turns on at night after having been off throughout the daylight hours. The only enablement hypotheses which can account for this delay are those in which the enablement path involves a *Thermal-Expansion* mechanism, with the initial cause being the cooling of the room. See Figure 5.23. Only this slow temperature change and the creeping motion it can produce can explain the observed delay. No "alarm clock" hypotheses are admitted, involving an *Electro-Mechanical* mechanism, with the initial cause being current at the outlet.

5.5.3 Abstractions and Shortcomings in the Home Heating Models

The proposed heat transport models for the home heating system expose some of the difficulties in representing and reasoning about mass quantities such as fluids. There are two complementary approaches to dealing with mass quantities in the literature: the *contained-stuff* and the *molecular-collection* paradigms [Hayes 79,85, Forbus 84,85, Collins and Forbus 87]. In the contained-stuff approach, instances of mass quantities are defined and distinguished by the containers which hold them. This paradigm supports rea-

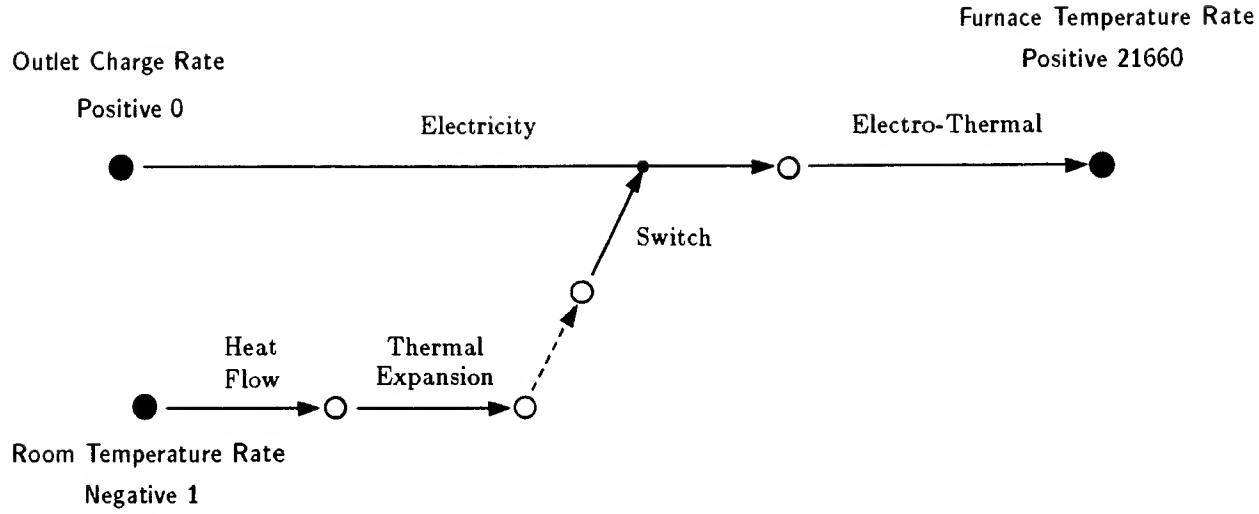


Figure 5.23. Thermostat hypothesis.

soning about global properties of mass quantities but not about how different parts of a contained mass quantity may participate in different mechanisms. The molecular collection approach, on the other hand, supports reasoning about how different parts of a mass quantity may be participating in different mechanisms at different sites and at different times, but not about interactions which involve an entire contained mass quantity.

A shortcoming of the pumped heat transport hypothesis for the home heating system is that the source of the *Fluid-Exchange* mechanism is not constrained to be coincident with the source of the *Heat-Flow* mechanism. In other words, the program JACK offers only the arrival of fluid at the radiator as an enablement explanation for heat flow from the furnace to the radiator, not the arrival of fluid *from the furnace* to the radiator.

The representations for the *Fluid-Exchange* and *Fluid-Heat-Transport* mechanisms follow the molecular collection paradigm. In the *Fluid-Exchange* mechanism, an amount change in a mass quantity at one site propagates to an amount change in a mass quantity at another site. In the *Fluid-Heat-Transport* mechanism, an amount change in a mass quantity at one site enables a temperature change at another site to propagate to a temperature change at the given site. This representation does not include mention of the

site from which the amount change in the mass quantity propagated.

A contained-stuff representation can partially overcome this shortcoming. Specifically, a fluid flow quantity can be associated with the *medium* {·RELATION· *Furnace Joined-To Radiator*} which spans the sites of the furnace and radiator. The medium represents the pipe which is a container; the fluid flow quantity represents global motion of a mass quantity associated with this medium. With this representation, the enablement path which terminates at the *Heat-Transport* mechanism involves fluid flow along the medium between the furnace and radiator, instead of only—and incompletely—an amount change in a mass quantity at the radiator. The form of this hypothesis is shown in Figure 5.24.

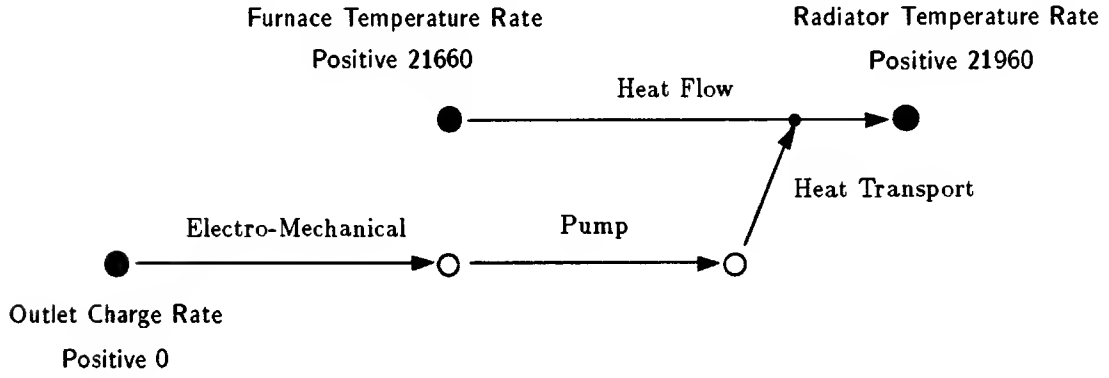


Figure 5.24. Contained-stuff heat transport hypothesis.

The contained stuff hypothesis shows how a fluid flow process can enable the *Heat-Flow* mechanism between the furnace and the radiator. In this hypothesis, the fluid flow process is subsumed into a single fluid flow quantity. Unfortunately, this subsumption confounds the separate source and destination of the fluid flow which are distinguished in the molecular collection representation.

A different approach to representing heat transport by a fluid involves buttressing the molecular collection representation by extending the ontology for causal graphs so that mechanisms, instead of just quantities, can enable other mechanisms. In particular, a *Fluid-Exchange* mechanism, in which both the fluid source at the furnace and the fluid destination at the radiator are

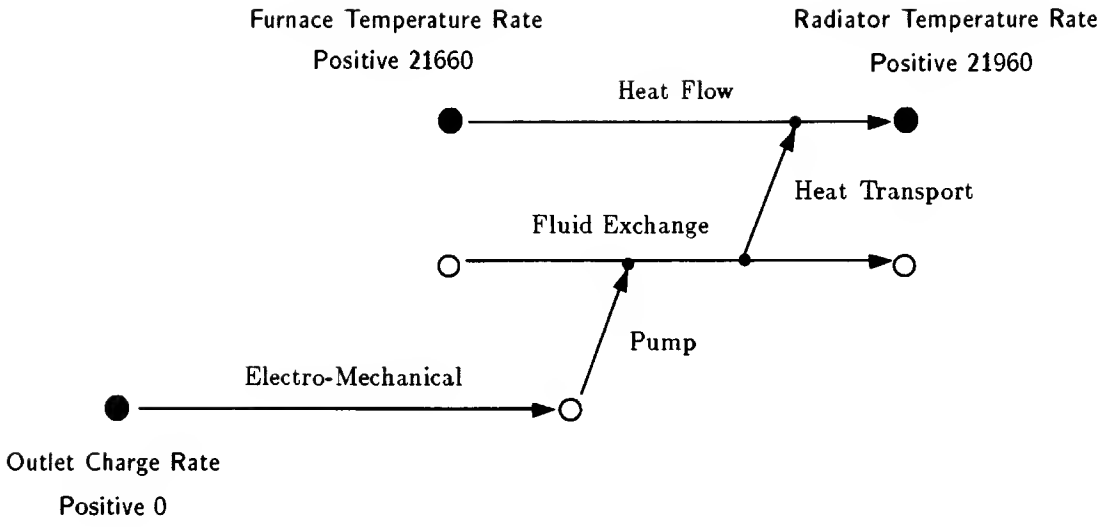


Figure 5.25. Molecular collection heat transport hypothesis.

represented, can enable the *Heat-Flow* mechanism via a *Heat-Transport* mechanism. The form of this hypothesis is shown in Figure 5.25.

In the contained-stuff hypothesis, the *Pump* mechanism does not enable a *Fluid-Exchange* mechanism in which an amount change in a mass quantity is propagated from one site to another; rather, the *Pump* mechanism directly causes a change in a flow quantity of a contained mass. This difference in the structure of the causal graph introduces a curious deficiency not present in the original fluid heat transport hypothesis: there is no hidden input. The fluid moving between the furnace and the radiator is represented as a singular contained stuff which can be directly acted upon by the *Pump* mechanism. In the molecular collection paradigm, where pieces of stuff are tracked explicitly from site to site, it is not possible to reason about a *Pump* mechanism without identifying a fluid source. In the original fluid heat transport hypothesis of Figure 5.21 there is a hidden input involving an *Amount-of-Fluid* quantity. In the extended molecular collection hypothesis of Figure 5.25 there are *two* hidden inputs: an *Amount-of-Fluid* quantity at the furnace and an *Amount-of-Fluid* quantity at the radiator. Furthermore, the *Fluid-Exchange* mechanism constrains these quantities to be of opposite sign. This hypothesis, and only this one, can be extended to include a cycle which describes the

closed circulation of fluid in a home heating system. This cycle hypothesis is shown in Figure 5.26.

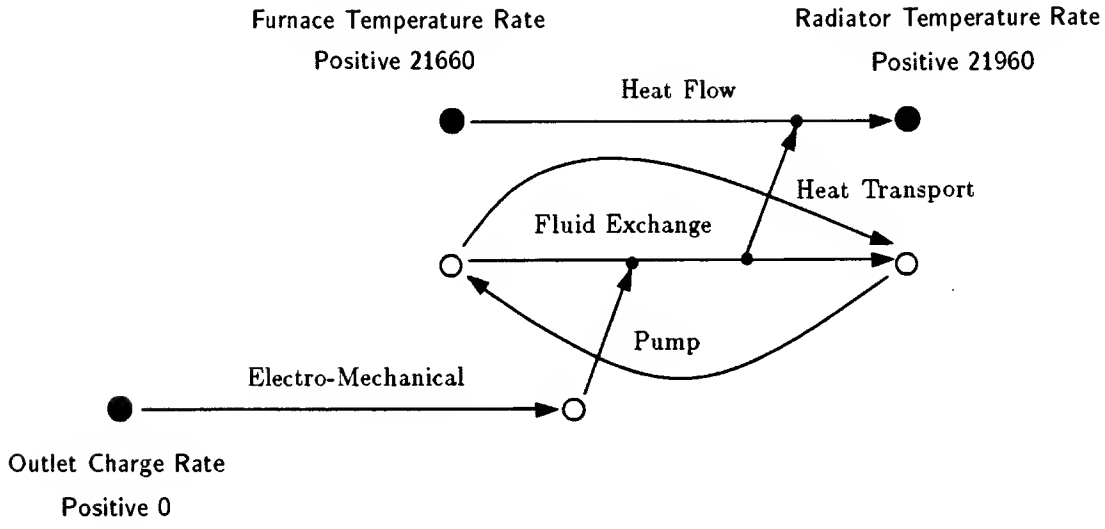


Figure 5.26. Heat transport cycle hypothesis.

This final model for a home heating system appears to combine the complementary advantages of the molecular collection and contained stuff paradigms for representing and reasoning about mass quantities.

6. Analysis of Results and Performance: JACK Be Simple, JACK Be Quick

My aims in this chapter are several: (1) to reveal the number of hypotheses generated by the program JACK for each of the device examples, (2) to separate the sources of pruning power which keeps the size of the hypothesis space manageable, (3) to examine the robustness of the approach in response to small changes in device observations and finally, (4) to outline assumptions and limitations in the approach.

6.1 Number of Hypotheses Admitted

The program JACK constructs causal graphs which connect observable events of a device. Each proposed causal graph is a possible explanation of some subset of the observable behavior of a device. From these causal explanation fragments, a complete and consistent model of a device can be built.

6.1.1 Grey Compartments

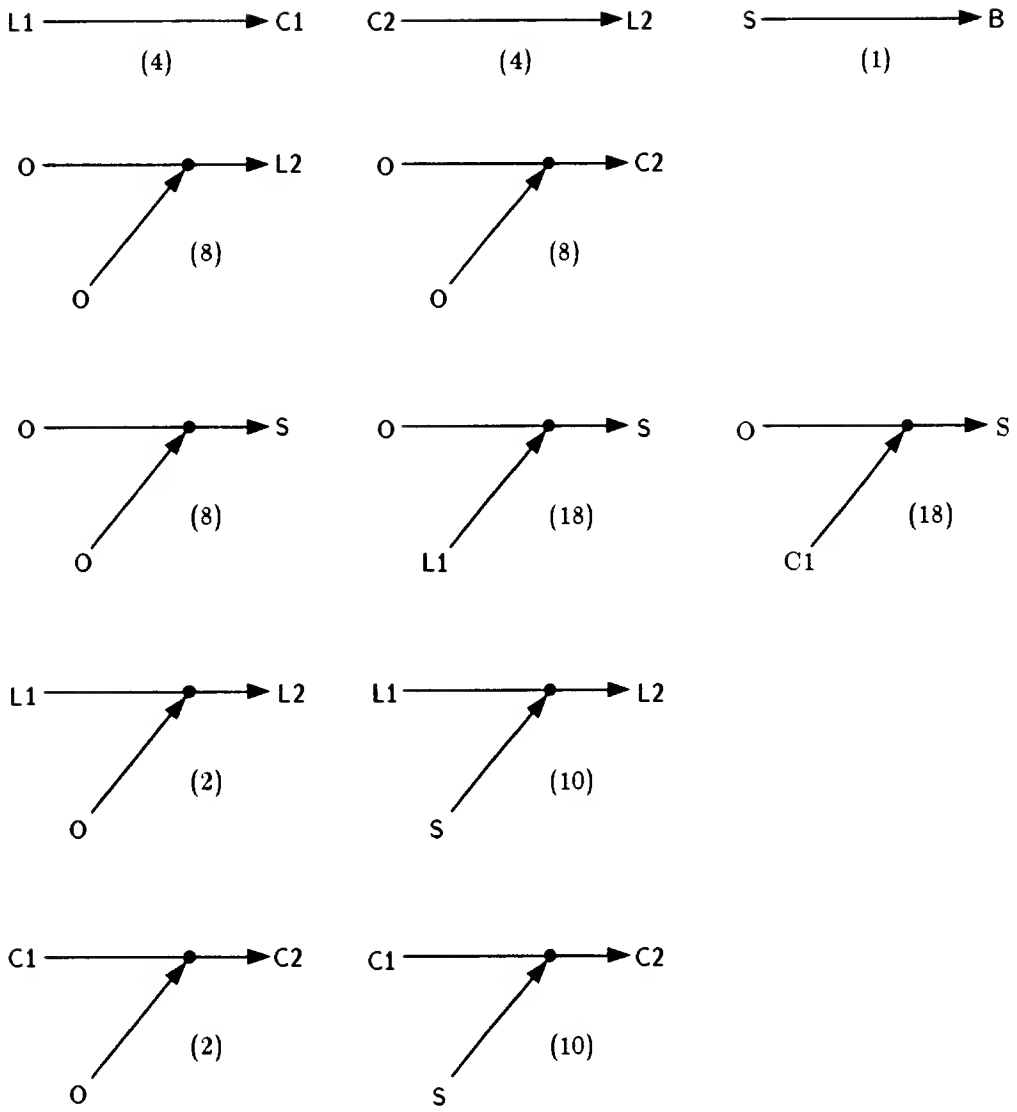
Each set of proposed causal graphs which connect the same subset of observable device events may be termed a "grey compartment"—grey because some light has been shed on the original black box in that specific sets of possible mechanism configurations have been enumerated; compartment because the proposed explanations cover some part of the black box of the entire device.

The grey compartments found by the program JACK for the toaster example are shown in Figure 6.1. The number of grey compartments found for each of the device examples is given in Table 6.1.

Device	*Grey Compartments*
Toaster	12
Tire Gauge	2
Bicycle Drive	3
Refrigerator	2
Home Heating	8

Table 6.1. Number of grey compartments.

Grey compartments form a useful abstraction space in which to reason about a device. The most important point is that the grey compartments



L1: Lever Position Rate Negative 60

L2: Lever Position Rate Negative 186

C1: Carriage Position Rate Negative 60

C2: Carriage Position Rate Negative 186

O: Outlet Charge Rate Positive 0

S: Coils Temperature Rate Positive 61

B: Bread Darkness Rate Positive 66

(n) number of causal graphs in grey compartment

Figure 6.1. Grey compartments for the toaster.

are decoupled from one another. Each grey compartment represents a set of possible causal explanations for the events of some set of *observable* device quantities. The viability of any grey compartment as a component in a complete model of a device depends only on whether at least one of the proposed mechanism configurations within a grey compartment continues to be consistent with subsequent observable events involving the same specific and no other quantities.

The usefulness of the abstraction space provided by grey compartments lies in this mutual independence of the compartments. Should all the proposed causal graphs within a grey compartment become refuted, all complete models built on that grey compartment also become refuted. For example, the thermal latch and motorized latch hypotheses for the toaster (see Section 5.1.2) correspond to different grey compartments. When the motorized latch compartment is refuted, all complete models of the toaster which incorporate one of the motorized latch hypotheses become unviable.

Complete models of a device are built from the grey compartments by conducting a straightforward graph search from the declared inputs of a device to its declared outputs. Starting from grey compartments whose causes correspond to device inputs, grey compartments are chained together matching effects to causes until all device outputs are reached. These complete models also are guaranteed to be consistent because of the mutual decoupling of the grey compartments. A complete and consistent model for the toaster in the grey compartment abstraction space is shown in Figure 6.2.

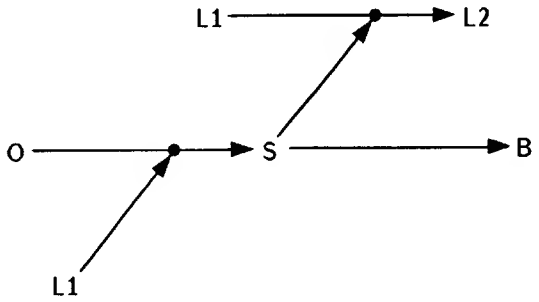


Figure 6.2. A complete and consistent model for the toaster.

Some care must be taken in indexing proposed causal graphs into grey compartments so that redundant grey compartments are not generated. In par-

ticular, the causal modelling system should notice when a conjectured hidden event in a proposed causal graph matches an observable event. See Figure 6.3. In this case, the proposed grey compartment is redundant because smaller grey compartments which subsume it have already been proposed. The observable events of a device impose a grain scale on the hypothesizing activity of the causal modelling system.

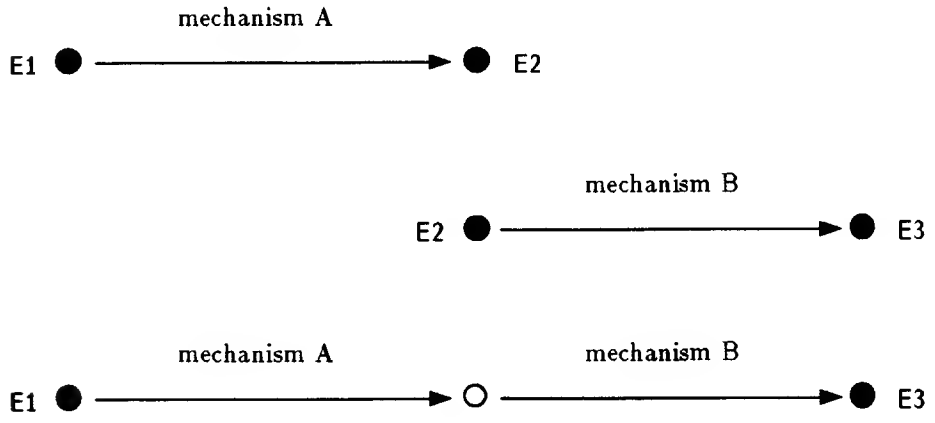


Figure 6.3. Redundant grey compartments.

There is a special case in which different grey compartments can be determined to be *a priori* mutually inconsistent. This is the case where the cause event and effect event of one grey compartment are juxtaposed in another grey compartment. An event cannot be both the cause and the effect of another event. Note that this ambiguity can occur only between simultaneous events.

This confusion arises in the toaster example in the simultaneous downward and upward motions of the lever and carriage. Does the lever move the carriage or vice versa? The ambiguity is resolved by teleological knowledge: the downward motion of the lever is declared to be a device input and cannot be an effect; the upward motion of the lever is declared to be a device output and cannot be a cause. In the case where neither event is a device input or output, the ambiguity can be resolved by determining which event is causally closer to a device input or output in a complete model of the device.

6.1.2 Causal Graphs

Each grey compartment is grey because there are several causal graphs which explain the observable events associated with the compartment. There remains some ambiguity concerning what mechanism configurations give rise to the parts of the observable behavior of a device defined by the grey compartments.

Table 6.2 shows the total number of causal graphs within the grey compartments admitted by the program JACK for each of the device examples. l_{max} is the length of the longest mechanism path in any causal graph for the given device. p_{max} is the greatest number of interacting mechanism paths in any causal graph for the given device. Refinement of hypotheses over multiple observations was disabled in these runs; the concern here is to determine the size of the initial set of causal graph hypotheses produced.

Device	l_{max}	p_{max}	*Causal Graphs*
Toaster	2	2	93
Tire Gauge	4	2	400
Bicycle Drive	2	2	31
Refrigerator	4	2	222
Home Heating	3	3	464

Table 6.2. Number of admitted causal graphs.

The number of causal graphs generated by the program JACK is reported as the sum of the causal graphs in the individual grey compartments to emphasize the fact that hypotheses in different grey compartments are decoupled from one another. Although it is true that, strictly speaking, the number of complete models for a device is a product and not a sum (it is the product of the numbers of causal graphs in the grey compartments chained together to form composite graphs which connect device inputs to device outputs), this much greater number is misleading. At no time is it desirable, sensible, or necessary to enumerate the complete causal models. When individual causal graphs or entire grey compartments are refuted, huge swaths of the space of complete causal models are pruned away. The grey compartments are decoupled by observable events, and the causal graph hypotheses within them can be constructed and refined in isolation from one another.

The overall amount of pruning achieved by the causal modelling system is impressive. For example, in the case of the tire gauge, the worst case number of hypotheses is on the order of $48^{4\times 2} \approx 10^{13}$.

In addition, the overall pruning ratio increases as the worst case number of hypotheses increases. Although not conclusive. (necessary but not sufficient) this property suggests that the performance of the causal modelling procedure is out of the exponential realm. The overall pruning ratio ranges from $\approx 10^5$ in the case of the toaster to $\approx 10^{11}$ in the case of the refrigerator.

6.2 Pruning Power

For each of the device examples, a potentially exponential hypothesis space is pruned down impressively. In this section, I examine the origins of this pruning power and assess the relative computational utility of the various sources of constraint in my approach to the causal modelling problem.

6.2.1 The Constraints

One of the sources of pruning power in the causal modelling system is the set of constraints on type, behavior, and structure. Table 6.3 shows the results of a set of experiments designed to isolate the pruning contributions of the various constraints. All of these experiments were run on the toaster example, with $l_{max} = 2$ and $p_{max} = 1$. The setting for p_{max} prevents hypothesizing from moving past the linear mechanism path level so that the pruning contribution of the hypothesis ordering is not confounded with that of the constraints.

The experiments are divided into two series. In one series—aimed at exposing the isolated pruning power of the individual constraints—the program JACK was run with a single constraint active. In the other series—aimed at exposing the incremental pruning power of the individual constraints—the program JACK was run a single constraint inactive. In addition, there were two control runs with all and none of the constraints active.

Constraint	*Pruning ratio when only constraint active*	*Pruning ratio when only constraint inactive*
Type	995	15.7
Delay	6.28	3640
Sign	3.04	7200
Direction	1.98	8190
Magnitude	4.77	5160
Alignment	3.10	7530
Bias	2.93	7060
Displacement	2.31	8190
Medium	2.31	8190
none	1.00	8190

Table 6.3. Relative pruning power of the constraints.

The type constraint clearly is the single most powerful constraint. After these experiments were completed, I modified the control structure of the causal modelling system to employ the type constraint as an initial generator: causal graphs which satisfy the type constraint are constructed in one phase, then the other constraints are applied in parallel.

However, the type constraint does not do all the work. The size of the hypothesis space after application of the type constraint remains too large to be termed manageable. For example, the number of hypotheses admitted by the type constraint alone in the tire gauge example is 5842. The other constraints further reduce the size of this set to 400.

To take the anthropomorphic viewpoint for a moment, the type constraint separates out the patently ridiculous hypotheses; the other constraints embody principles which allow finer distinctions to be made among the remaining hypotheses.

A rough understanding of the unique power of the type constraint can be gained by considering the amount of focusing which occurs as values are propagated across mechanisms for the various constraints. I define focusing here as the ratio of the number of values propagated to the number of possible values. The number of values propagated for the type constraint across any number of mechanisms is always one; if the number of distinct types is n_t, the focusing ratio is always $1/n_t$. For the current vocabulary of mechanisms, $n_t = 9$.

The focusing ratios for the constraints whose values are qualitative regions never can be as favorable. The number of possible values is fewer and all

possible values can be generated across as few as a single mechanism. In particular, all possible values can be generated for the direction and alignment constraints after propagation across a single mechanism. For the sign and bias constraints, the highly ambiguous {*Negative Positive*} value set can be generated across a single mechanism. Finally, for the displacement constraint, all possible values can be generated across a mechanism path of length two.

A brief inspection of Table 6.3 reveals that the pruning attributable to the delay and magnitude constraints, whose propagated values are ranges of orders of magnitude, is approximately twice that of the constraints based on qualitative regions, yet still falls well short of that of the type constraint.

The number of possible values for the delay and magnitude constraints is 36, ranging from 2^{-17}—taken to be the value for zero, to 2^{18}—taken to be the value for infinity. Despite the greater number of possible values, the focusing ratios for these constraints remain poor because the propagated order of magnitude ranges can expand quickly to encompass the entire set of contiguous possible values. The origin of the weak focusing is the often wide order of magnitude ranges which specify the time constants and efficiencies of mechanisms.

The values propagated for the medium constraint are the subjects and objects of structural relations. Knowledge of structural relations within a device is almost always lacking given the "black box" nature of the modelling problem. In the absence of this knowledge, unbound physical objects are propagated. Of course, these conjectured physical objects are compatible with any observed effect event. The pruning power of the medium constraint in the absence of structural knowledge can never be high because unbound variables represent any number of possible values and manifest the poorest possible focusing ratio of 1.0.

A perusal of the second column of Table 6.3 reveals that the marginal pruning contribution of the weaker constraints approaches zero. In these particular runs disabling the direction, displacement, and medium constraints made no difference in the size of the hypothesis set. These empirical results suggest that additional physical and causal constraints may not bear much fruit. Constraint sources of a different character are needed.

6.2.2 The Ordering on Hypotheses

The constraints on type, behavior, and structure reflect physical and causal principles which are used to test hypotheses. The hypothesis ordering, on the other hand, is used to control the generation of hypotheses. The aim of the ordering is to suppress the extension of hypotheses into more complex or

less constrained ones unless there is a compelling reason for suspecting that augmentation will result in more complete hypotheses.

Table 6.4 shows the pruning which takes place between levels of the hypothesis ordering. The layers of pruning are attributable to the heuristic recognition rules associated with the hypothesis levels. Linear mechanism path hypotheses which fail to satisfy all the constraints are passed through the mechanism interaction recognition rule. All hypotheses—even those which satisfy all the constraints—are passed through the hidden input recognition rule. Finally, pairs of hidden input hypotheses are passed through the cycle recognition rule.

Hypothesis Level	*Pruning Ratio at Level*
Mechanism Interactions	149
Hidden Inputs	5.46
Cycles	2.30

Table 6.4. Pruning power of the recognition rules.

The greatest amount of pruning achieved across a level jump is the pruning associated with the recognition rules which justify the jump from the linear mechanism path level to the mechanism interaction level. The importance of focusing between these two levels cannot be understated for the worst case number of hypotheses increases here from exponential in one parameter to exponential in two parameters. The mechanism interaction recognition rules are second only to the type constraint as a source of pruning power.

The importance of the pruning which takes place between the linear mechanism path level and the mechanism interaction level is apparent in the tire gauge example. Only 18 of 2663 failed hypotheses satisfied the mechanism interaction recognition rules. These 18 hypotheses were augmented into 385 admitted disablement and equilibrium interaction hypotheses. Without the pruning due to the recognition rules for mechanism interactions, the explosion here would have been truly prohibitive.

The pruning ratio associated with the jump to the hidden input hypothesis level directly reflects the frequency of occurrence of the "interaction-only" mechanisms along conjectured mechanism paths. The low pruning ratio during hypothesis generation at this level is unfortunate because hidden inputs compromise the ability to test hypotheses by comparing predicted events to observed events. The type constraint is the only constraint of any utility once the assumption that all inputs are observable is removed.

As expected, hidden input hypothesis construction is mostly successful

because there is no observable event against which to compare inferred events. For example, in the case of the home heating system, 1390 triggerings of the hidden input recognition rule led to 382 admitted hidden input hypotheses. Most of the admitted hypotheses for the home heating system are hidden input hypotheses.

The recognition rule for cycles identifies pairs of hidden input hypotheses made up of one conjectured source and one conjectured sink. The pruning ratio associated with this recognition rule approaches the limiting case where the number of hypotheses involving hidden sources and the number of hypotheses involving hidden sinks are equal. In the limiting case, exactly half of the possible pairs satisfy the rule.

Cycle hypothesis construction also has a high success rate. In the refrigerator example, 9966 pairings of hidden input hypotheses resulted in 2010 successful cycle hypotheses. Failures mostly are due to sources and sinks being of incompatible types, being associated with different physical objects, and having magnitude ranges which do not overlap so that they do not cancel when added together.

6.2.3 Abstraction by Type

Abstraction spaces based on type provide yet another means of controlling search in causal modelling. The idea behind abstraction spaces is to partition a set of primitives into a hierarchy of classes and to perform search in stages. Initial search at the coarse level of the classes is used to focus search at the finer grain size of the primitives.

The primitives in causal modelling are mechanisms. Type, which already has proven to be the single most powerful source of constraint, is used as the basis for partitioning mechanisms into classes. Mechanisms which map the same cause type into the same effect type are collected into the same class. For example, the *Rigid-Coupling* and *Forward-Ratchet* mechanisms, both of which map {·TYPE· *Position Rate*} into {·TYPE· *Position Rate*}, are collected into the class *Mechanical-Coupling*.

The composite mechanisms which represent classes are formed by combining the descriptions of the constituent mechanisms. More specifically, the union of the qualitative region sets specified in the constituent mechanisms is taken for the sign, direction, alignment, bias, and displacement constraints. The union of the order of magnitude ranges specified in the constituent mechanisms is taken for the delay and magnitude constraints. And the union of the structural relations specified in the constituent mechanisms is taken for

the medium constraint. The mechanism classes formed in this way are shown in Appendix B.

Table 6.5 indicates the search reduction achieved in the abstract mechanism space for each of the device examples. The results of these less costly searches can be used to focus search at the level of the primitive mechanisms by substituting constituent mechanisms for composite mechanisms in the hypotheses generated in the abstraction space. These more detailed hypotheses are tested by propagating constraints in the usual manner.

Device	*Causal Graphs w/o Type Abstraction*	*Causal Graphs w/ Type Abstraction*
Toaster	93	27
Tire Gauge	400	48
Bicycle Drive	31	4
Refrigerator	222	60
Home Heating	464	190

Table 6.5. Search reduction in type abstraction space.

Another way of utilizing the mechanism classes stems from the observation that the classes are nearly closed under composition. Stated differently, the constraints on behavior and structure imposed by any mechanism path of length two or greater comprised solely of mechanisms from a single class are almost always already represented in a single mechanism of the same class. For example, a *Rigid-Coupling* mechanism composed with a *Contact-Coupling* mechanism is indistinguishable from the *Contact-Coupling* mechanism. Some mechanisms are not composable at all: for example, the *Forward-Ratchet* and *Backward-Ratchet* mechanisms.

Closure under composition within a mechanism class does not hold categorically, however. For example, the efficiency of a chain of *Conductive-Heat-Flow* mechanisms is lower than a single such mechanism. Nevertheless, this property of mechanism classes forms the basis of a heuristic which provides another source of pruning power. This heuristic states that adjacent mechanisms in a mechanism path cannot be of the same class.

6.3 Robustness

The device observations input to the causal modelling system are idealizations of what real perceptual data might be. In any investigation where idealized data is substituted for real data, there always is a possibility that too much

of the desired output is encoded in the input. In partial response to this potential charge, I point to the alternate hypotheses admitted by the causal modelling system for each of the device examples.

In order to ascertain more fully whether I had inadvertently and subtly guided the hypothesizing activity of the program JACK by encoding too much of the target models in the device observations, I conducted experiments to determine the robustness of the causal modelling system's performance in the face of changes in the device observations. These experiments fall into two categories: those where spurious detail is added to a device observation and those where the "black box" is opened partially and pertinent detail is added to a device observation.

6.3.1 More Irrelevant Detail

The experiments in which irrelevant detail was added to a device observation took the following form: The quantity {·QUANTITY· *Earth Gravity Amount*} was added to both the toaster and the tire gauge observations. Motions form part of the observable behavior of both of these devices; nevertheless, gravity plays no role in the operation of either.

The causal modelling system is misled in the case of the toaster: a hypothesis involving the *Gravity* mechanism is admitted as a possible explanation for the observed downward motion of the carriage.

In the case of the tire gauge, a *Gravity* hypothesis fails to explain the observed motion of the slide. The direction constraint is violated because the orientation of the slide motion is not downward. This basis for pruning the hypothesis is not particularly convincing; one can easily imagine a hypothesis which includes a mechanism which alters the direction of the motion. However, the magnitude constraint provides a more compelling justification for pruning this hypothesis: the rate at which the slide moves is too high to be due to gravity.

The causal modelling system exhibited acceptable performance in these instances of irrelevant events being included in device observations. Some spurious events were incorporated into admitted hypotheses; in other cases, the physical and causal principles embedded in the constraints provided the basis for removing hypotheses concerning these events from consideration.

6.3.2 More Relevant Detail

The experiments in which relevant detail was added to a device observation took the following form: The structural relations {·RELATION· *Cylinder*

Contains Piston}, {·RELATION· *Piston Attached-To Slide*}, {·RELATION· *Piston Connected-To Slide*}, and {·RELATION· *Piston Touches Slide*} were added to the tire gauge observation.

No truth-value histories were associated with these relations; they served only to declare the existence of physical objects and structural pathways inside the device. They had the effect of constraining the causal modelling system to incorporate declared physical objects in hypotheses rather than to freely conjecture hidden physical objects.

Care was taken to avoid predisposition towards particular mechanisms. Specifically, the several declared relations involving the *Piston* and the *Slide* constrain the causal modelling system to incorporate the previously unknown *Piston* in hypotheses without introducing a preference for any particular mechanism of the *Mechanical-Coupling* class.

With this ancillary knowledge of the internal structure of the tire gauge available, one would expect that—via the medium constraint—a number of additional hypotheses could be eliminated. This indeed proved to be the case, as shown in Table 6.6. The experiment was conducted with the vocabulary of primitive mechanisms and in the abstraction space.

Type Abstraction?	*Causal Graphs w/o structural knowledge*	*Causal Graphs w/ structural knowledge*
no	400	202
yes	48	17

Table 6.6. Performance with internal structural knowledge.

6.4 Assumptions and Limitations

My purpose in this section is to root out assumptions embedded in my approach to the causal modelling problem and to delineate limits in that approach on generating and distinguishing hypotheses about mechanisms within devices.

6.4.1 Closed-World Assumptions

Closed-world assumptions are impossible to avoid; nevertheless one must strive to be aware of them and understand their impact. In my work, closed-world assumptions manifest in the vocabulary of mechanisms and at the ontological level. Mechanisms such as pulleys and friction and magnetism, to name

a few, do not appear in the vocabulary. Furthermore, certain types of causal graphs do not appear in the ordering on hypotheses. These include iterative cycles and interactions wherein entire mechanisms (as opposed to quantities) enable or disable other mechanisms. As a result of these closed-world assumptions, hypotheses concerning friction in the tire gauge and concerning fluid flow enabling heat flow in the home heating system are excluded from consideration.

6.4.2 Hidden Parameters

Much of the potential pruning power of the constraints based on ranges of orders of magnitude—the delay and magnitude constraints—is compromised by hidden parameters in the mechanism descriptions. For example, the wide range specified for the time constant of the *Fluid-Flow* mechanism reflects uncertainty about the unknown path length between source and destination. Similarly, the wide range of efficiency specified in the *Electro-Thermal* mechanism reflects uncertainty about the unknown resistance of the material.

The uncertainty associated with hidden parameters is endemic, for they correspond to unobservable quantities. There is little point in making these suppressed unobservable quantities explicit, for this just pushes the uncertainty "over the horizon" without reducing it.

6.4.3 Tradeoff Interactions

The only type of mechanism interaction in the ordering on hypotheses involving additive contributions is the equilibrium interaction. There is no tradeoff interaction, in which unbalanced additive contributions result in a net change in the direction of the largest contribution.

The most obvious signature for a tradeoff interaction is a non-zero effect of an unaccountably reduced magnitude. Unfortunately, a recognition heuristic based on this signature proves ineffective, because the order of magnitude ranges propagated for the magnitude constraint typically are wide enough to account for too-low magnitudes arising from tradeoff interactions.

An alternate recognition heuristic is based on the observation that a hypothesis identifying only the "background" contribution of a tradeoff interaction fails to account for the observed direction of change. The way to repair such an incomplete hypothesis is to conjecture an additional, dominant contribution in the opposite direction. At this time, a tradeoff interaction recognition heuristic has not been implemented.

6.4.4 Higher-Order Derivatives

Second and higher-order derivatives can be represented by multiple instances of the temporal integration mechanism along mechanism paths. However, higher-order derivatives in individual mechanisms—such as acceleration due to gravity—are not represented explicitly. Instead, a range of efficiency is associated with the mechanism which approximates the results of integration. For example, in the gravity mechanism, this efficiency range results in a range of values being propagated for the velocity associated with the effect. This treatment of higher-order derivatives is admittedly rough because the interval over which temporal integration occurs is not represented explicitly but is hard-coded in the efficiency range.

6.4.5 Monotonicity and Linearity

The dependencies between quantities due to mechanisms are assumed to be both monotonic and linear. These assumptions are embedded in the procedures for refining hypotheses. For example, once the sense of a mechanism dependence has been established in one observation, it is assumed, by monotonicity, to be the same in other observations. The monotonicity assumption may result in the unjustified retraction of a hypothesis.

The impact of the linearity assumption on hypothesis refinement is more subtle. Without knowledge of the order of a dependence, no attempt can be made to determine if observed magnitude differences across multiple examples of behavior are consistent with linearity or non-linearity. This potential basis for retracting hypotheses is unavailable.

6.4.6 Representation of Structure

The representation for structure encoded in the displacement and medium constraints supports descriptions only of distinguishable physical objects and simple physical connections between objects. This fairly impoverished representation for structure results in abstract models of devices whose physical structure is complex: to wit, the models of the latch mechanism in the toaster and the brake mechanism in the bicycle drive. Some ideas concerning additional aspects of externally observable structure which can place constraints on internal mechanism configurations are presented in Section 7.4.3.

6.4.7 Teleology

The teleological reasoning exhibited by the program JACK is limited. Only

one form of teleological knowledge is provided in the device observations: some events are labelled as known inputs or known outputs. The known inputs cannot appear as effect events in any hypothesis and the known outputs cannot appear as cause events.

Only one of the recognition rules which control passage between levels in the ordering on hypotheses has a teleological basis: the cycle recognition rule is based on the principle that synergistic cycles can remove potential sources and sinks. This is fundamentally a rule of design; models with the extra sources and sinks are physically realizable and explain the same observations.

In this work, I have focused on the the use of causal reasoning in the modelling problem. I expect fully that teleological knowledge would prove to be a complementary source of pruning power. However, by making a clean separation between causal and teleological reasoning, the approach to modelling described in this thesis can in principle be applied to domains where teleological knowledge is lacking, such as natural systems, and domains where teleological knowledge is inarticulate, such as economics.

7. Lessons Learned: The Morals of the Story

In this final chapter, I articulate the principles behind the reasoning exhibited by the causal modelling system in generating and distinguishing hypotheses, I reexamine the set of issues set out at the beginning of the thesis, I compare the results of my research effort to those achieved in other related efforts, I outline some future directions for my work, including some scenarios on how causal models constructed by the program JACK can support problem solving tasks, and finally, I offer some ideas concerning possible applications for a causal modelling system.

7.1 Principles

All research efforts in artificial intelligence must be evaluated on two criteria: the generality of the set of principles articulated in the work, and the computational utility of those principles. In Chapter 6, I analyzed the performance of the causal modelling system JACK. In this section, I enumerate the principles which underlie my approach to causal modelling. The diverse reasoning exhibited by the causal modelling system on the set of implemented device examples serves as a demonstration of the generality of these principles. These principles fall into two categories: (1) the constraints on type, behavior, and structure and (2) the rules for recognizing incomplete hypotheses at different levels of the ordering on hypotheses.

7.1.1 Physical and Causal Constraints

All of the constraints support reasoning about how mechanisms map device inputs to device outputs. Each constraint concerns a different observable aspect of the behavior and structure of physical systems. All hypotheses about hidden mechanism configurations within devices must account for any observed changes or lack of changes between cause events and effect events for all of these aspects of behavior and structure.

The type constraint concerns the types of quantities in physical systems. Hypotheses must account for observed type conservations or transformations between causes and effects. The type constraint turns out to be the single most powerful constraint.

Several conservation laws from physics are mingled in this constraint. The law of conservation of energy which permits the form of energy to change while it is conserved is reflected in type transformations. The law of conservation of mass is reflected in type conservations which concern mass quantities such as

fluids. The law of conservation of momentum is reflected in type conservations which concern motion.

The delay constraint concerns the times of occurrence of events in physical systems. Hypotheses must account for observed time lags between causes and effects. The delay constraint figures prominently in the generation of enablement, disablement, and equilibrium hypotheses. A reliable hallmark of unsuspected mechanism interactions is an unexplained delay.

The delay constraint embodies the causal principle that effects cannot precede their causes and the physical principle that the propagation time for any interaction must be finite. Perception plays a limiting role, however. The distances to be crossed may be so short (the extent of a single physical object) and the speeds of propagation may be so high (the speed of light) that causally connected events may be perceived to be simultaneous.

The sign constraint concerns the signs of the values of quantities in physical systems. Hypotheses must account for any change or lack of change of sign between causes and effects. The sign constraint plays an important role in the construction of equilibrium hypotheses in the tire gauge example and cycle hypotheses in the refrigerator example. In all of these hypotheses, the interacting additive contributions must be of opposite sign.

The sign constraint captures the sense, direct or inverse, of the dependencies between quantities due to mechanisms. In addition, when applied to mechanisms of flow, the sign constraint reflects the essence of the concept of conservation: a substance can be redistributed within a system, but can never be consumed or spontaneously created.

The direction constraint concerns the orientations of quantities in physical systems. Hypotheses must account for any deflections between causes and effects. The direction constraint supports reasoning about the behavior of springs in the toaster and tire gauge examples.

The direction constraint embodies both Newton's law of inertia and the law of conservation of momentum. The descriptions of mechanisms which concern motion state explicitly whether or not the direction of motion may be altered between cause and effect. Furthermore, deflections notwithstanding, motions must be transferred; they cannot vanish between cause and effect.

The magnitude constraint concerns the magnitudes of the values of quantities in physical systems. Hypotheses must account for any decreases, increases, or lack of change in magnitude between causes and effects. The magnitude constraint plays the key role in distinguishing thermal expansion as the mechanism employed in the thermostat of the home heating system.

The magnitude constraint reflects two physical principles: the law of conservation of energy, and the notion of mechanical advantage. Energy can

never be lost, but the efficiency of energy transfer across a mechanism may be less than perfect. Also, it is possible to achieve amplifications across a mechanism, albeit always at the expense of some other quantity. For example, a small gear driven by a large gear will revolve faster, the speed of a fluid in a pipe may be increased by constricting the diameter of the pipe, etc.

The alignment constraint concerns the relative values of quantities in physical systems. Hypotheses must incorporate any inequality relations imposed by mechanisms between causes and effects. The alignment constraint is indispensable in revealing the one-way nature of the linkages in the tire gauge and bicycle drive examples.

The bias constraint concerns the directions of change of quantities in physical systems. Hypotheses must incorporate any restrictions concerning absolute directions of change imposed by mechanisms between causes and effects. The bias constraint supports reasoning about the complementary roles of condensation and evaporation in the refrigerator example.

Two physical principles concerning directionality are reflected in the alignment constraint: The thermodynamical principle of entropy, and the notion of "path of least resistance" which states that motion or flow is always towards points of locally lesser potential. The bias constraint reflects other origins of directionality in asymmetrical processes and mechanisms, such as those due to geometry.

The displacement constraint concerns the locations of objects in physical systems. Hypotheses must account for any physical separation between causes and effects. The displacement constraint plays a role in reasoning about how a moving fluid or gas can transport heat in the home heating system.

The medium constraint concerns the structural connections between objects in physical systems. Only those hypotheses for which the appropriate structural connections between causes and effects can be established or conjectured may be admitted. The medium constraint supports reasoning about collapsing equilibrium states within the tire gauge.

The displacement and medium constraints reflect the causal notion of no action at a distance. Mechanisms must span any separation between cause and effect. And more specifically, causal interactions can take place only if the appropriate structural connections are established.

7.1.2 Rules for Traversing the Hypothesis Ordering

The ordering on hypotheses is a manifestation of Occam's Razor. Hypothesis types are organized into a hierarchy according to two simplicity met-

rics: the worst-case number of possible hypotheses for each type and the potential for constraining hypotheses of each type by observable events.

The simplest control structure for traversing the hypothesis ordering involves exhaustively generating hypotheses at each level and proceeding to the next level only when all hypotheses at the previous level have been eliminated. There are two problems with this form of control: (1) It is likely that many observations are needed before all hypotheses at a given level can be eliminated. My intuition is that hypotheses generation from first principles can and should proceed further with limited observations. (2) More importantly, no focusing occurs as control is passed to the more complex or less constrained hypothesis levels. The explosion lurking at the next level is merely delayed but not contained.

We want a control structure which supports justified and focused excursions into the more complex or less constrained hypothesis levels. Control never should be passed indiscriminately between levels. There should be always a clear justification for moving from one level of hypothesis construction to another, and there should be always a sharp focusing to offset the potential explosion. The control structure I have implemented supports justified and focused jumps between levels in the hypothesis ordering and is based on the following principle:

Incomplete hypotheses exhibit characteristic deficiencies.

Manifestations of deficiency justify attempts to augment hypotheses at another level of hypothesis construction. Focusing occurs in two ways: The deficiency signatures indicate into what other form of hypothesis a hypothesis should be extended. Furthermore, only certain deficient hypotheses ever are augmented; in particular, no admitted hypothesis ever is extended.

For each level in the hypothesis ordering beyond the root level, there is a rule used to recognize situations in which an attempt to augment a deficient hypothesis at that level of hypothesis construction might be successful. Each rule is a different instantiation of the principle that incomplete hypotheses exhibit characteristic deficiencies.

Unsuspected mechanism interactions manifest in some combination of unexplained delays, magnitudes, and signs. Unexplained delays manifest because the time of occurrence of an interaction is always the later of the times of occurrence of the contributions. Stated differently, both contributions must occur before an interaction occurs.

Unexplained magnitudes manifest in enablement situations because an effect may be disrupted over a continuous range. A fluid rate controlled by a valve may take on a range of values, whereas electric current controlled by a switch is either on or off.

Unexplained signs and magnitudes manifest in disablement and equilibrium situations because the non-zero effects expected from the contributions of single mechanism paths are always at variance with the zero effects which result from disablement and equilibrium interactions.

Some mechanisms can only enable and disable other mechanisms. These mechanisms have no power to produce effects other than through interaction with other mechanisms. For example, a switch can disrupt electrical flow, but cannot produce current in the absence of a current source. Similarly, a pressure change can result in a heat loss through the process of evaporation, but only in the presence of a heat sink.

Hidden inputs are implied whenever "interaction-only" mechanisms are hypothesized outside of interactions. These incomplete hypotheses may be accompanied by unexplained signs and magnitudes because isolated "interaction-only" mechanisms cannot account for non-zero effects. The missing inputs which can explain observed non-zero effects may be present, even though candidates for these inputs have not been identified among observable events.

The recognition rule for cycles, unlike the other recognition rules, is based on a principle of design: well-designed physical systems have a minimum of sources and sinks. Potential sources and sinks within a device may be avoided by forming cycles where gains alternate with losses so that both are always temporary and bounded. The possibility for such a synergistic cycle exists whenever both an undeclared source and an undeclared sink are proposed as part of the explanation of the behavior of a physical system.

7.2 The Issues Revisited

In this section, I relate the degree of success achieved in addressing the set of issues outlined in Chapter 1.

How to constrain the formation of hypotheses?

With a set of physical and causal constraints and an ordering on hypotheses.

My approach to making the causal modelling problem tractable is a two-pronged approach. One of the prongs involves applying a set of constraints which embody physical and causal principles to prune hypotheses. The other prong involves enumerating different forms for hypotheses, placing an ordering on these forms, and using this ordering to carefully control the generation of hypotheses. The pruning power resulting from the combined application of these thrusts has proven to be impressive.

What are the constraints in the physical system domain?

Among them are type, delay, sign, magnitude, direction, alignment, bias, displacement, and medium.

I have identified a useful, although certainly not exhaustive set of constraints for reasoning in the physical system domain. These constraints highlight diverse aspects of the behavior and structure of devices. They provide several dimensions along which to reason, including: the types of quantities, the times, locations, and magnitudes of events, orientations in space, the causal pathways between events, and restrictions on directions of change due to dependencies between quantities, the direction of causation, and one-way behavior. These constraints are derived from physical and causal principles and represent necessary conditions which all physically realized devices must satisfy.

What are the different causal structures for devices?

Among them are linear mechanism paths, mechanism interactions including enablements, disablements, and equilibria, hidden inputs, and cycles.

I have enumerated a number of different causal structures for devices. These include simple linear mechanism chains from inputs to outputs, enablement, disablement, and equilibrium interactions between mechanisms where multiple causes combine into single effects, and synergistic cycles where gains and losses within a device offset each other. Any of these hypothesis forms for devices may involve primitive causes or initial inputs which are not observable.

How to deal with the complexity vs. completeness problem?

By utilizing the principle that incomplete hypotheses often exhibit characteristic deficiencies.

I have placed on ordering on the different hypothesis forms for devices based on a straightforward complexity analysis of the corresponding causal structures. A heuristic rule is associated with each level in the hypothesis ordering. These rules are used to recognize characteristic deficiencies in hypotheses and justify attempts at augmentation at more complex or less constrained levels of the hypothesis ordering.

The ordering serves as a manifestation of Occam's razor: the simplest hypothesis forms are considered first; hypothesis forms deeper in the ordering are considered only when simpler forms exhibit signs of incompleteness. Within each level, hypotheses are generated in order of the length of mechanism paths and the number of mechanism interactions. All hypotheses within each level of the ordering can be generated and the ordering is extensible.

What is the power of causal reasoning in the mechanical, electrical, and thermal domain?

Comparable to its power in the circuit domain, given a uniform representation for describing diverse mechanisms.

One of the results of this research effort is a uniform representation for describing a wide variety of mechanisms in devices. This representation supports reasoning about mechanical, electrical, and thermal phenomena in the toaster, mechanical and pneumatic phenomena in the tire gauge, mechanical phenomena in the bicycle drive, mechanical, electrical, and thermal phenomena in the refrigerator, and mechanical, thermal and hydraulic phenomena in the home heating system.

This uniform representation and the set of procedures which operate on it support reasoning about causal relations between events due to hypothesized mechanisms. This causal reasoning is at the core of my solution to the "black box" problem for devices.

What makes for a convincing model of a device?

Abstractions which preserve physical plausibility, predictive power, and the ability to support problem solving tasks.

Abstraction is an important aspect of the causal modelling process. Detailed reasoning about hidden mechanism configurations within devices is infeasible due to the size of the hypothesis space. In addition, precise reasoning is impossible due to lack of knowledge about unobservable mechanisms and events.

My goal in designing the constraints on type, behavior, and structure was to capture compact sources of discriminatory power strong enough to overcome the unavoidable dearth of information due to the "black box". These constraints are based on principles from physics and causality. They reflect conditions which all designs for devices must satisfy. Ideally, we want a causal modelling system which removes physically implausible hypotheses from consideration and which is able to make fine distinctions among physically plausible hypotheses. Results from the several implemented device examples indicate that the causal modelling system does not admit physically implausible hypotheses. These same results offer many instances of successful discrimination among physically plausible hypotheses.

For example, the program JACK is able to distinguish toaster models in which a latch on a spring is released either through thermal expansion or a motor, is able to expose the one-way nature of the coupling between a hidden piston and the slide in a tire gauge, and is able to determine that an evaporation/condensation model for a refrigerator can include a cycle but an evaporation/space heater model cannot. Moreover, the toaster models are distinguished when a prediction derived from the thermal latch hypothesis is compatible with an additional observation while a prediction derived from the motorized latch hypothesis is not.

Remaining confusions among hypotheses are often traceable to the loss of resolution arising from default values in mechanism descriptions. In other cases, conceivable distinctions cannot be described within the given set of constraints or the given ontology of causal graphs. In still other cases, competing hypotheses represent genuine, if abstract, alternate designs for devices. For example, a refrigerator could be designed with a renewable fluid input to seed evaporation.

In Section 7.4.1, I offer scenarios in which causal models constructed by the program JACK support diagnosis and monitoring tasks.

7.3 Relation to Other Work

In this section, I discuss a number of themes of current interest in causal and qualitative reasoning in the context of my work and the work of others. In addition, I look at a different research project concerning theory formation for devices. Finally, I interpret causal modelling as an instance of Waltz network labelling.

7.3.1 Causal and Qualitative Reasoning

Many approaches to causal and qualitative reasoning have appeared in the literature. Seminal works among these include Forbus' Qualitative Process Theory [Forbus 84], de Kleer and Brown's qualitative physics based on confluences [de Kleer and Brown 84], and Kuipers' method for inferring behavior from causal structure [Kuipers 84].

These and other research efforts all identify composability as an important property of causal reasoning: the overall behavior of a physical system must be derivable from its topology and the behavior of its constituents. The composability of the mechanism descriptions in the vocabulary of the causal modelling system derives from the uniform representation for the constraints on type, behavior, and structure and the propagation and combination procedures which operate on this representation.

The same research efforts indicate that causal and qualitative reasoning subsumes several complementary forms of inference. There are techniques for reasoning about dynamics—which changes occur?, time—when do events occur?, physical objects—where do events occur?, topology—what are the causal pathways?, thresholds—what new values are reached?, and preconditions—which mechanisms are active and which are inactive? Here I describe how the procedures in the program JACK support all of these forms of reasoning. Where appropriate, I compare my inference method to others in the literature.

The type, delay, sign, magnitude, and medium constraints collectively determine which changes occur in a physical system. The medium constraint addresses the question of where events occur by determining the physical objects involved. Physical objects along with types determine the quantities involved. The delay constraint addresses the question of when do events occur by determining moments. The sign and magnitude constraints jointly determine values. Inferred quantities, moments, and values collectively determine events, which addresses the question of which changes occur.

The medium constraint directly addresses the question of what are the causal pathways by determining structural connections between the sites of events. My treatment of causal pathways improves on others described in the literature. Kuipers does not make structural connections explicit; his "structure" is actually the set of functional relationships or qualitative differential equations which describe a physical system. de Kleer and Brown do not allow for time-dependent structural connections. In my approach, structural connections are represented by relations with histories, enabling the program JACK to verify the existence of causal pathways at different times. Furthermore, the causal modelling system can reason about how changing quantity values establish or disrupt causal pathways. In enablement hypotheses, the medium of the enabled mechanism is affirmed at the moment of the interaction; in disablement hypotheses, the medium of the disabled mechanism is denied at the moment of the interaction. These assertions make for a compact description of active and inactive mechanisms, useful in generating predictions in the context of additional observations. Although Forbus does allow for the declaration of causal pathways in process descriptions, they are treated as a static input; there is no reasoning about the interplay between quantity values and physical structure.

The temporal integration procedure in the program JACK addresses the question of what new values are reached. The temporal integration schemes employed by Forbus, de Kleer and Brown, and Kuipers use only the direction of change and the value space to determine the next value of a quantity. My scheme also employs the magnitude of the rate and the duration of change. Although there is certainly nothing new here, this more complete rendering of temporal integration allows finer distinctions to be made. For example, the "alarm-clock" and "thermostat" hypotheses for the toaster are distinguished by determining that a higher initial temperature in the toaster implies a shorter interval until a latch on a spring is released.

Kuipers' causal and qualitative simulation system QSIM is able to reason about undeclared landmark values of quantities. Similarly, the causal modelling system JACK is able to construct disablement and equilibrium hypotheses

about observed stable values which are not declared limit values in the value spaces of quantities.

The question of which mechanisms are active and which are inactive is addressed by several of the constraints. Preconditions concerning threshold values of quantities are handled by the sign and magnitude constraints and the temporal integration procedure. Preconditions concerning the presence or absence of structural connections are handled by the medium constraint. Preconditions concerning relative values of quantities, for example that the pushing object must be behind the pushed object in a contact coupling, are handled by the alignment constraint. Preconditions concerning absolute directions of change, as in a ratchet, are handled by the bias constraint.

The "no-function-in-structure" principle of de Kleer and Brown states that individual mechanism descriptions must not be able to explain behavior which arises from interactions with other mechanisms. This principle is reflected in the handling of hidden input hypotheses in the causal modelling system. A *Switch* mechanism alone cannot account for a current flow; a current source must also be identified. Similarly, an *Evaporation* mechanism cannot explain cooling in the absence of a heat sink.

The "no-function-in-structure" principle and its dual "no-structure-in-function," are reflected also in the clean separation in the program JACK of the constraints which concern behavior (delay, sign, direction, magnitude, alignment, and bias) from those which concern structure (displacement and medium). These constraints are mutually independent; no value propagated for any constraint depends on a value propagated for any other.

In his work on diagnostic reasoning based on structure and behavior [Davis 84], Davis enlists a hypothesis ordering to control search through a hypothesis space. Each level in his ordering on fault hypotheses corresponds to the removal of a different assumption about how devices are supposed to work. The ordering relations are derived not from a complexity analysis, but from empirical knowledge concerning the frequency of different kinds of faults. Hypothesis orderings, whatever their derivation, are an effective and general way of dealing with the complexity vs. completeness problem.

Another theme in Davis' work is the use of multiple representations for structure, each highlighting a different kind of causal pathway in devices and each corresponding to a different manifestation of the concept of adjacency. In my work, the several constraints serve as multiple representations, supporting reasoning about physical systems along several different dimensions of type, behavior, and structure.

The utility of multiple representations is self-evident: a distinction which is muddled in one representation can be sharp in another. The key idea again

is abstraction. Which details to expose and which to suppress? The set of constraints in the causal modelling system provide an assortment of compact sources of discriminatory power. Any one of the constraints can provide the key to distinguishing hypotheses. A difficult issue in the use of multiple representations is how to select or even construct the right representation for the task at hand. I have addressed this issue partially in my work. Choices about which form of hypothesis to consider are supported by the recognition rules for the levels in the hypothesis ordering. However, no choices are made concerning which subset of the constraints to apply to a given modelling task.

Dependencies between quantities are assumed to be linear (see Section 6.4.5) in my approach to causal modelling. This qualitative approach to reasoning about the behavior of physical systems is limited in that rates of change are abstracted and interesting properties such as extrema, stability and asymptotic approach cannot be represented [Kuipers 85]. Sacks' piecewise linear reasoning approach [Sacks 87,88] offers a useful compromise: higher-order equations are approximated by linear segments, retaining some of the information (e.g., maxima and minima) lost in the qualitative approach.

The procedures for refining hypotheses in the program JACK embody a simple form of comparative analysis. These procedures support predictions about activations and deactivations of mechanisms and changes in delays and magnitudes across different observations. Weld has provided a comprehensive treatment of comparative analysis [Weld 88]. He offers two complementary methods, differential qualitative analysis and exaggeration, which together solve many problems in predicting how a device responds to perturbations of its parameters.

7.2.2 Theory Formation for Devices

Surprisingly enough, at least to me, there is a dearth of other work on theory formation for devices. The notable exception is Shrager's work on a theory of human instructionless learning [Shrager 87]. Shrager describes a method called view application whereby device hypotheses are incrementally refined by incorporating abstract schemas into developing models. One of Shrager's explicit goals is psychological validity. My approach concentrates on the sources of constraint and the causal reasoning which make the modelling problem tractable. Our approaches are complementary.

7.2.3 Waltz Labelling

Causal modelling can be cast as an instance of Waltz network labelling

[Waltz 75]. The networks are the causal graphs which represent hypotheses about hidden configurations of mechanisms within devices. The arcs are labelled with mechanisms and the nodes are labelled with values for the constraints which describe events.

A singular difference separates the causal modelling problem from other instances of Waltz labelling—the network is not known. Only the peripheral nodes and their labellings are known. These are the externally observable events. During the causal modelling process, networks are constructed by conjecturing mechanism arcs and additional event nodes.

The performance of the program JACK offers an extraordinarily convincing demonstration of the potential power of the Waltz network labelling technique. In the right domain and with the right constraints, the network need not even be known. Network topologies can be generated in concert with the actual labelling process, and this process can still result in a hypothesis set of manageable size.

7.4 Future Work

In this section, I present scenarios which illustrate how causal models produced by the program JACK can be used in problem solving, and I explore ideas about how to further limit search in causal modelling, how teleological reasoning can be incorporated into causal modelling, and the role which experiment design can play in causal modelling.

7.4.1 Using Causal Models

The acid test for a device model is whether or not it can support problem solving. An important form of problem solving in the physical system domain is diagnosis. Here I present a scenario which shows how causal models of a tire gauge constructed by the program JACK can be used in troubleshooting a misbehaving tire gauge.

Imagine a tire gauge is giving an incorrect pressure reading. In particular, imagine this reading is too high. Consider the "spring" (see Figure 5.8) and "impulse" (see Figure 5.9) hypotheses generated by the causal modelling system. Further consider types of failure in which the value space of a quantity becomes shifted or truncated so that one of the limit values is unattainable. See Figure 7.1. This kind of behavior might manifest when a device component is broken, or bent, or becomes displaced.

In the spring hypothesis, either type of fault would result in the equilibrium state which halts the motion of the slide being achieved at a different

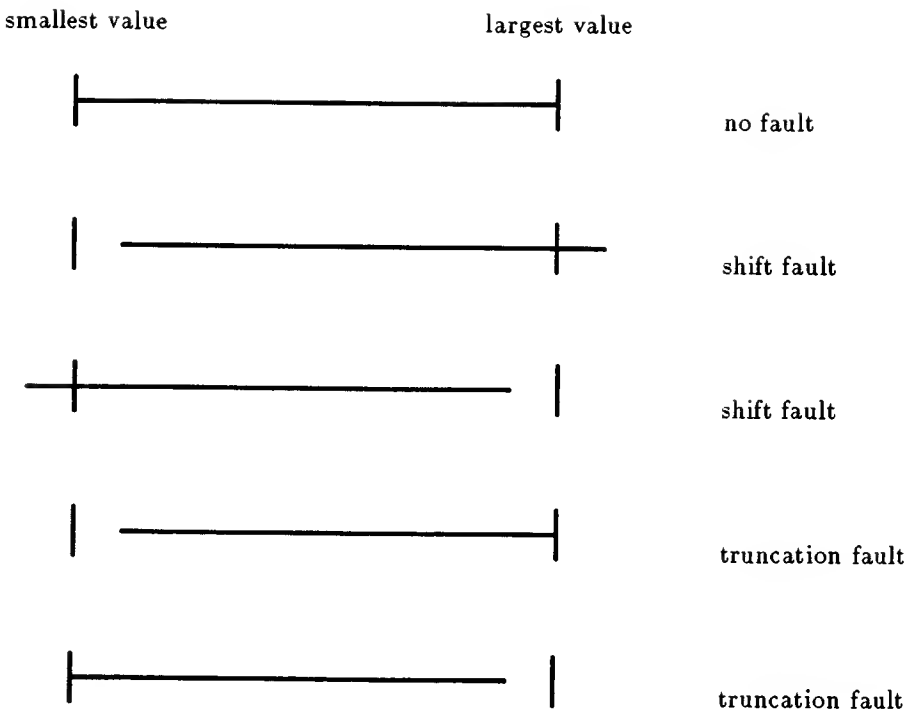


Figure 7.1. Shifted and truncated value spaces.

position. In particular, a final position of the slide corresponding to a higher pressure reading is possible.

In the impulse hypothesis, neither type of fault can explain a higher pressure reading. In the cases where the disabling (smallest, by convention) value cannot be attained, the motion of the slide would not cease and it would move all the way to its limit of motion. In the cases where the range of motion of the valve has been shifted or truncated towards the disabling value, the valve would close sooner than in the nominally operating device, resulting in a lower pressure reading.

Another problem solving task in the physical system domain is monitoring—verifying the nominal operation of a device. Causal models can support efficient, reliable device monitoring without the need for exhaustive checking of all observable quantities.

A causal model reveals causal dependencies among events in a physical

system. An analysis of these causal dependencies can support decisions about which events to monitor. In particular, the importance of events can be assessed by determining how many other events are effects or causes of a given event. The importance of an event is related to the amount of subsequent activity it supports, and the amount of activity which arranges for its occurrence. Events which lie on more than one mechanism path should be verified with care. On the other hand, events which are side effects and do not support further activity of the device need be given only cursory attention, if at all.

Figure 7.2 shows a causal model for a toaster generated by the program JACK. In this figure, next to each observable event is the number of events which are causes or effects of the given event. By this analysis, the upward and downward motions of the lever and the heating of the coils are the most informative events in the toaster. The other observable events cannot as reliably verify that the toaster is working properly.

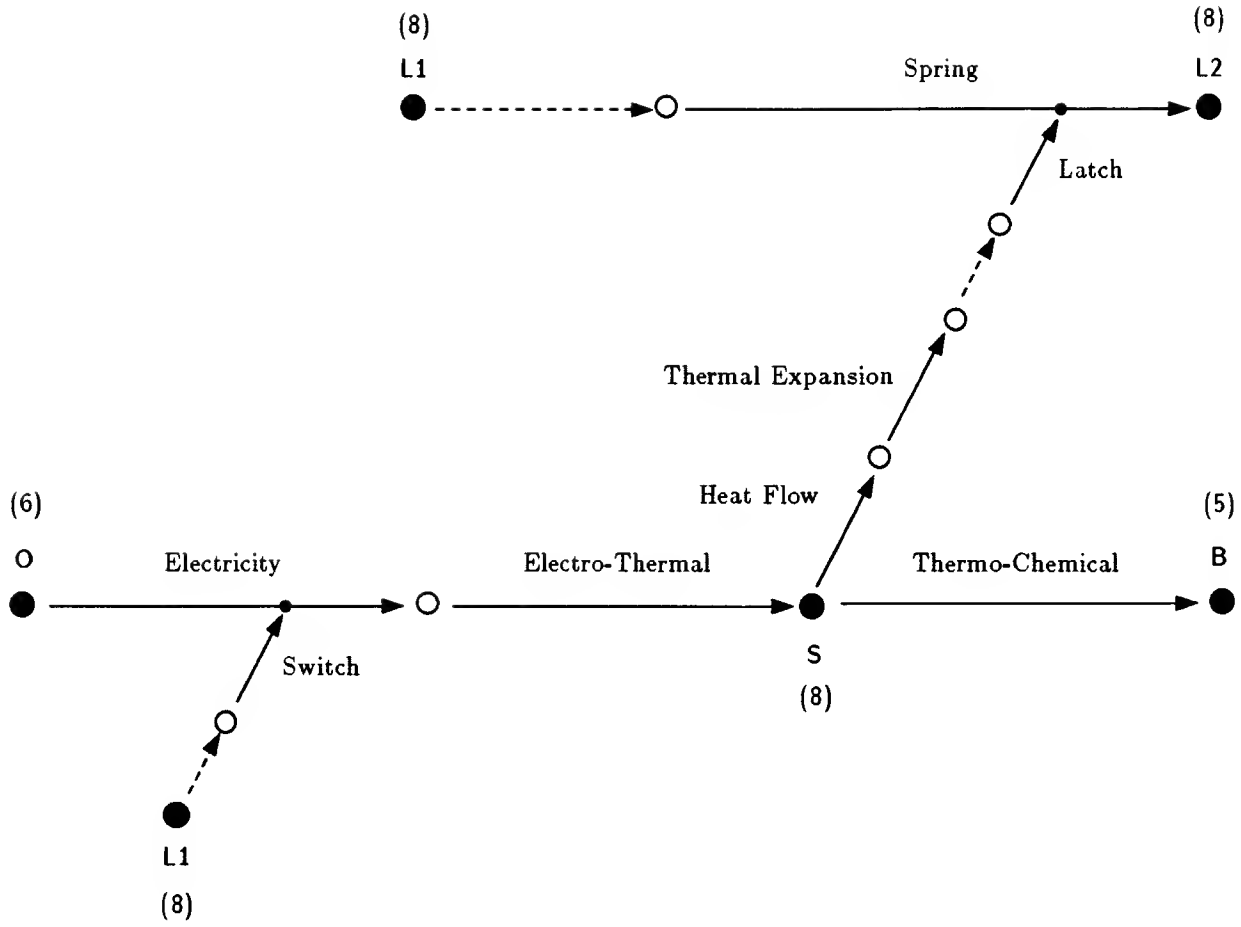
In Appendix F, these ideas on using causal models in device monitoring are elaborated further. There I describe the transfer of results from my thesis work to a project at the Jet Propulsion Laboratory.

7.4.2 Limiting Search

Differential diagnosis is a technique used in medical diagnosis to limit the size of the set of hypotheses [Patil et al 82, Pople 82]. The technique is a straightforward manifestation of beam search. Generated hypotheses are ordered according to a set of criteria and a cutoff threshold is established. Clinical tests then are selected to distinguish only these best hypotheses.

The abstraction space of mechanisms used to focus search in the program JACK was formed by grouping into classes mechanisms which map the same cause type into the same effect type. There are many other bases for forming *a-kind-of* hierarchies for mechanisms: mechanisms which span two locations vs. those which take place within a single physical object, mechanisms which conserve type vs. those which transform type, mechanisms which impose constraints on directions of change vs. those which do not, etc.

Each a-kind-of hierarchy reflects different choices about which distinctions to expose and which to suppress. Ideally, the selection and/or construction of abstraction spaces should be sensitive to the modelling task at hand. Given an observation, which sources of discrimination are likely to prove most useful in distinguishing hypotheses to explain the observed events? For example, are there many quantities which change in one direction but not the other? The issue of choosing among multiple abstraction spaces indicates an intriguing direction in which to extend this work. Of relevance is Lathrop's



L1: Lever Position Rate Negative 60

L2: Lever Position Rate Negative 186

O: Outlet Charge Rate Positive 0

S: Coils Temperature Rate Positive 61

B: Bread Darkness Rate Positive 66

(n) number of causes and effects of event

Figure 7.2. Monitoring a toaster.

work on developing methods for constructing abstractions in several domains, including the physical system domain [Lathrop et al 87a, Lathrop et al 87b, Lathrop 88].

Another way to limit search is to coarsen the grain of the search space by combining primitives into macros. Macros reduce search by highlighting some subset of the possible ways to compose primitives. Good candidates for macros in causal modelling are those mechanism compositions which made multiple appearances in the device models generated by the program JACK. An example is the "thermostat" hypothesis in which motion due to a *Thermal-Expansion* mechanism results in an enablement.

Explanation-based learning methods [Mitchell et al 86, DeJong and Mooney 86] are techniques for forming macros from those compositions of primitives which contribute to successful problem-solving episodes. The idea is to use experience as the filter to decide which macros are worth forming and using. This form of learning could be used to incrementally enhance the performance of the program JACK. An early effort on my part to explore the use of explanation-based learning in causal modelling is described in [Doyle 86].

7.4.3 Teleological Reasoning

The teleological reasoning in the program JACK is limited: Events declared as known inputs cannot be effects and events declared as known outputs cannot be causes. Cycle hypotheses do not explain any additional behavior; instead they reflect a principle of design which states that the number of inputs to a device should be reduced whenever possible.

Other forms of teleological knowledge can help to constrain hypothesizing—for example, declarations concerning *intended* causal dependencies in a device. In a refrigerator, the cooling of the interior is intended, the warming of the exterior is not. This distinction can offer a different kind of clue for the presence of a cycle. In a typical synergistic cycle, one half of the cycle lies on a causal path leading to intended outputs, the other half exists only to remove the potential source or sink associated with the first half. This is true of the cycle in a refrigerator. The evaporation half of the cycle powers the intended cooling of the interior. The warming of the exterior arising from the condensation half of the cycle is a side effect.

Still other sources of discriminatory power can be gleaned from the simple fact that a device is a *designed* artifact. A designer always satisfies constraints other than the inviolable ones due to physics and causality. Designs also reflect constraints of a pragmatic nature which provide different dimensions along which to reason about a device. Among the possibilities are: cost—

is the set of conjectured mechanisms consistent with the price tag of the device; availability—do any of the proposed mechanisms involve materials not readily obtainable; size and layout—can the hypothesized configuration of mechanisms be packed into the observed volume of the device; and weight—is the set of conjectured mechanisms consistent with the heaviness of the device.

7.4.4 Experiment Design

Designing experiments is premature while the number of possible hypotheses is overwhelming. The causal modelling system produces manageably sized sets of hypotheses about the mechanisms in devices by reasoning from first principles. The program JACK also can refine hypotheses over multiple observations of a device. A clearly indicated next step is to introduce an experiment design capability for determining how best to alter the configuration of a device to actively and efficiently distinguish hypotheses.

Jonathan Amsterdam is doing exactly that. He has begun an investigation into the role of experiment design in theory formation, using the causal models output by the program JACK as a starting point.

Other relevant work includes de Kleer and Williams' minimum entropy method for determining the site of the most discriminating next observation of a circuit [de Kleer and Williams 87], Shirley's work on efficiently generating test vectors in troubleshooting [Shirley 86], and the work of Rajamoney and others on a general experiment design capability [Rajamoney et al 85, Rajamoney and DeJong 87].

7.5 Applications for a Causal Modelling System

In this final section, I discuss two possible practical applications in the long term for a causal modelling system.

7.5.1 Early Design

Engineers make use of numerous abstractions when they first tackle design tasks. The set of constraints in the causal modelling system capture a particular set of abstractions for reasoning about devices. These constraints support the kind of rough, "within-an-order-of-magnitude" reasoning typical of the early stages of the design of a device. A causal modelling system could be used to generate a set of abstract, physically plausible designs for a device. The engineer could then proceed to the more difficult task of tweaking and tuning first-cut designs until specifications are met.

A causal modelling system could also permit the engineer to explore designs for a device without having to enumerate formal specifications. The designer could construct "observations" which instantiate the desired behavior of the device and could slowly converge on the set of specifications needed in the later stages of design.

Ulrich is developing a similar approach to the conceptual design of electromechanical systems [Ulrich 88]. He describes a method for specifying schematic descriptions of devices which meet a behavioral specification. The approach involves generating rough solutions to meet nominal input-output behavioral specifications and then debugging these prototype designs to meet full behavioral specifications. His design and debug approach helps to harness the potential combinatorial explosion of possible solutions to design problems. Ulrich's representations for mechanisms are based on bondgraphs [Rosenberg and Karnopp 1983].

7.5.2 Modelling In-Line with Problem Solving

Causal models support numerous problem solving tasks concerning devices: verifying the nominal operation of a physical system, diagnosing faults, planning how to use a device to achieve particular goals. Models are inevitably incomplete for they are always constructed in the context of particular problem solving tasks. An automated causal modelling capability presents an intriguing possibility—the on-demand augmentation of models in the face of deficiencies exposed in the context of new problem solving tasks. A causal modelling system might generate hypotheses to explain alarm situations in a nuclear power plant, or might extend a model of a camera to support reasoning about strategies for taking fast-action sports photographs.

Admittedly, the results of this thesis represent a modest step towards automated modelling, but such a capability would have clear impact both inside and outside the academic community. Researchers could reduce their overhead by utilizing off-the-shelf domains. Even more importantly, these domains could be standardized; new research results would be accepted more quickly into a growing corpus. In the pragmatic world of knowledge-based expert systems, automatic modelling could help circumvent the knowledge engineering bottleneck. Furthermore, knowledge-based systems could have their knowledge bases extended as needed, rather than failing gracelessly in the context of new problem solving tasks.

References

[Allen 83] Allen, James, "Maintaining Knowledge About Temporal Intervals," *Communications of the ACM*, **26**, 1983.

[Barrow 84] Barrow, Harry G., "VERIFY: A Program for Proving the Correctness of Digital Hardware Designs," *Artificial Intelligence*, **24**, 1984.

[Collins and Forbus 87] Collins, John W. and Kenneth D. Forbus, "Reasoning About Fluids via Molecular Collections," *National Conference on Artificial Intelligence*, Seattle, 1987.

[Davis 84] Davis, Randall, "Diagnostic Reasoning Based on Structure and Behavior," *Artificial Intelligence*, **24**, 1984.

[Dean and McDermott 87] Dean, Thomas, and Drew V. McDermott, "Temporal Data Base Management," *Artificial Intelligence*, **32**, 1987.

[DeJong and Mooney 86] DeJong, Gerald and Raymond Mooney, "Explanation-Based Learning: An Alternative View," *Machine Learning*, **1**, no. 2, 1986.

[de Kleer 84] de Kleer, Johan, "How Circuits Work," *Artificial Intelligence*, **24**, 1984.

[de Kleer and Brown 84] de Kleer, Johan and John S. Brown, "A Qualitative Physics Based on Confluences," *Artificial Intelligence*, **24**, 1984.

[de Kleer and Williams 87] de Kleer, Johan and Brian C. Williams, "Diagnosing Multiple Faults," *Artificial Intelligence*, **32**, 1987.

[Doyle 86] Doyle, Richard J., "Constructing and Refining Causal Explanations from an Inconsistent Domain Theory," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

[Doyle et al 86] Doyle, Richard J., David J. Atkinson, and Rajkumar S. Doshi, "Generating Perception Requests and Expectations to Verify the Execution of Plans," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

[Doyle et al 87] Doyle, Richard J., Suzannne M. Sellers, and David J. Atkinson, "Predictive Monitoring Based on Causal Simulation," *Second NASA Artificial Intelligence Forum*, Palo Alto, California, 1987.

[Forbus 84] Forbus, Kenneth D., "Qualitative Process Theory," *Artificial Intelligence*, **24**, 1984.

[Forbus 85] Forbus, Kenneth D., "The Problem of Existence," *Cognitive Science Society*, Irvine. California, 1985.

[Forbus 86] Forbus Kenneth D., "Interpreting Measurements of Physical Systems," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

[Fox and Smith 84] Fox, Mark S. and Stephen Smith, "The Role of Intelligent Reactive Processing in Production Management," *CAM-I, 13th Annual Meeting and Technical Conference*, St. Paul. Minnesota, 1984.

[Genesereth 84] Genesereth. Michael R.. "The Use of Design Descriptions in Automated Diagnosis," *Artificial Intelligence*, **24**, 1984.

[Gini et al 85] Gini, Maria, Rajkumar Doshi, Sharon Garber. Marc Gluch, Richard Smith, and Imran Zualkernain, "Symbolic Reasoning as a Basis for Automatic Error Recovery in Robots," Technical Report 85-24, University of Minnesota. 1985.

[Hayes 79] Hayes, Patrick J.. "The Naive Physics Manifesto," in *Expert Systems in the Micro-Electronic Age*, D. Michie (ed.), Edinburgh University Press, 1979.

[Hayes 85] Hayes, Patrick J., "Naive Physics I: Ontology for Liquids," in *Formal Theories of the Commonsense World*, J. Hobbs and B. Moore (eds.), Ablex Publishing Corporation, 1985.

[Kuipers 84] Kuipers, Benjamin P., "Commonsense Reasoning About Causality: Deriving Behavior from Structure," *Artificial Intelligence*, **24**, 1984.

[Kuipers 85] Kuipers, Benjamin P., "The Limits of Qualitative Simulation," *Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, 1985.

[Lathrop et al 87a] Lathrop, Richard L.. Teresa A. Webster. and Temple F. Smith, "ARIADNE: Pattern-Directed Inference and Hierarchical Abstraction in Protein Structure Recognition," *Communications of the Association for Computing Machinery*, **30**, no. 11, 1987.

[Lathrop et al 87b] Lathrop, Richard L.. Robert J. Hall. and Robert S. Kirk. "Functional Abstraction from Structure in VLSI Simulation Models," *24th IEEE/ACM Design Automation Conference*. Miami Beach. Florida. 1987.

[Lathrop 88] Lathrop. Richard L.. *An Algorithm for Learning to Construct Abstractions*, Ph.D. Thesis, Department of Electrical Engineering and

Computer Science, Massachusetts Institute of Technology, *in progress* 1988.

[Mackie 74] Mackie, John L., *The Cement of The Universe: A Study of Causation*, Oxford University Press, 1974.

[Marr 82] Marr, David, *Vision*, W. H. Freeman and Company, 1982.

[McAllester 80] McAllester, David A., "The Use of Equality in Deduction and Knowledge Representation," Technical Report 550, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1980.

[Miller 85] Miller, David P., *Planning by Search through Simulations*, Ph.D. Thesis, Department of Computer Science, Yale University, 1985.

[Mitchell et al 86] Mitchell, Tom, M., Richard M. Keller, and Smadar T. Kedar-Cabelli, "Explanation-Based Generalization: A Unifying View," *Machine Learning*, **1**, no. 1, 1986.

[Patil et al 82] Patil, Ramesh S., Peter Szolovits, and William B. Schwartz, "Modeling Knowledge of the Patient in Acid-Base and Electrolyte Disorders," in *Artificial Intelligence in Medicine*, P. Szolovits (ed.), Westview Press, 1982.

[Pople 82] Pople, Harry E., Jr., "Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics," in *Artificial Intelligence in Medicine*, P. Szolovits (ed.), Westview Press, 1982.

[Rajamoney et al 85] Rajamoney, Shankar, Gerald F. DeJong, and Boi Faltings, "Towards a Model of Conceptual Knowledge Acquisition through Directed Experimentation," *Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, 1985.

[Rajamoney and DeJong 87] Rajamoney, Shankar, and Gerald DeJong, "The Classification, Detection, and Handling of Imperfect Theory Problems," *Tenth International Joint Conference on Artificial Intelligence*, Milan, 1987.

[Rieger and Grinberg 77] Rieger, Chuck and Milt Greenberg, "The Declarative Representation and Procedural Simulation of Causality in Physical Mechanisms," *Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, 1977.

[Rosenberg and Karnopp 1983] Rosenberg, Ronald C. and Dean C. Karnopp, *Introduction to Physical System Dynamics*, McGraw-Hill, 1983.

[Sacerdoti 77] Sacerdoti, Earl A., *A Structure for Plans and Behavior*, Elsevier North-Holland, 1977.

[Sacks 87] Sacks, Elisha P., "Piecewise Linear Reasoning," *National Conference on Artificial Intelligence*, Seattle, 1987.

[Sacks 88] Sacks, Elisha P., *Automatic Qualitative Analysis of Ordinary Differential Equations Using Piecewise Linear Approximations*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1988.

[Shirley 86] Shirley, Mark H., "Generating Tests by Exploiting Designed Behavior," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

[Shoham 86] Shoham, Yoav, "Chronological Ignorance: Time, Knowledge, Nonmonotonicity and Causation," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

[Shrager 87] Shrager, Jeffrey C., "Theory Change via View Application in Instructionless Learning," *Machine Learning*, **2**, no. 3, 1987.

[Ulrich 88] Ulrich, Karl T., *Computation and Pre-Parametric Design*, Sc.D. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, *in progress* 1988.

[Vere 83] Vere, Steven A., "Planning in Time: Windows and Durations for Activities and Goals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-5**, no. 3, 1983.

[Waltz 75] Waltz, David, "Understanding Line Drawings of Scenes with Shadows," in *The Psychology of Computer Vision*, P. Winston (ed.), McGraw-Hill, 1975.

[Weld 86] Weld, Daniel S., "The Use of Aggregation in Qualitative Simulation," *Artificial Intelligence*, **30**, 1986.

[Weld 88] Weld, Daniel S., *The Theory and Evaluation of Comparative Analysis Techniques*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1988.

[Williams 84] Williams, Brian C., "Qualitative Analysis of MOS Circuits," *Artificial Intelligence*, **24**, 1984.

[Williams 86] Williams, Brian C., "Doing Time: Putting Qualitative Reasoning on Firmer Ground," *National Conference on Artificial Intelligence*, Philadelphia, 1986.

How Things Work, **1-4** Edito-Service, S.A., Geneva.

Appendix A: The Device Observations

```
;;;;;THE TOASTER;;;;;

 ;;;PHYSICAL OBJECTS;;;
(define-physical-objects LEVER DIAL CARRIAGE COILS BREAD
  OUTLET EARTH)

 ;;;QUANTITIES;;;
(define-quantities
  (LEVER POSITION)
  (DIAL ANGLE)
  (CARRIAGE POSITION)
  (COILS TEMPERATURE)
  (BREAD APPEARANCE)
  (OUTLET CHARGE)
  (EARTH GRAVITY))

 ;;;QUANTITY SPACES;;;
(define-quantity-space LEVER POSITION Amount
  (DOWN -0.1) (UP 0.0))
(define-quantity-space LEVER POSITION Rate
  (NEGATIVE -0.1) (ZERO 0.0) (POSITIVE 0.1))

(define-quantity-space DIAL ANGLE Amount
  (L (* (// *pi* 6) 5)) (LM (* (// *pi* 3) 2))
  (M (// *pi* 2))
  (MD (// *pi* 3)) (D (// *pi* 6)))
(define-quantity-space DIAL ANGLE Rate
  (NEGATIVE -1) (ZERO 0) (POSITIVE 1))

(define-quantity-space CARRIAGE POSITION Amount
  (DOWN -0.1) (UP 0.0))
(define-quantity-space CARRIAGE POSITION Rate
  (NEGATIVE -0.1) (ZERO 0.0) (POSITIVE 0.1))

(define-quantity-space COILS TEMPERATURE Amount
  (OFF 30) (WARM 30 200) (HOT 200 300))
(define-quantity-space COILS TEMPERATURE Rate
  (NEGATIVE -1) (ZERO 0) (POSITIVE 2))

(define-quantity-space BREAD APPEARANCE Amount
  (UNTOASTED 0) (LIGHT 50) (GOLDEN 100)
  (MEDIUM 200) (BROWN 400)
```

```
 (DARK 800) (BURNT 2000))
(define-quantity-space BREAD APPEARANCE Rate
 (ZERO 0) (POSITIVE 1))

(define-quantity-space OUTLET CHARGE Amount
 (ON 100.0))
(define-quantity-space OUTLET CHARGE Rate
 (ZERO 0.0) (POSITIVE 10.0))

(define-quantity-space EARTH GRAVITY Amount
 (G 9.8))
(define-quantity-space EARTH GRAVITY Rate
 (ZERO 0.0))

 ;;;ZEROS;;;
(define-zero LEVER POSITION UP)
(define-zero DIAL ANGLE)
(define-zero CARRIAGE POSITION UP)
(define-zero COILS TEMPERATURE OFF)
(define-zero BREAD APPEARANCE UNTOASTED)
(define-zero OUTLET CHARGE)
(define-zero EARTH GRAVITY)

 ;;;DIRECTIONS;;;
(define-direction UP)
(define-direction DOWN)
(define-direction CLOCKWISE)
(define-direction SCALAR)

(define-quantity-direction UP (LEVER POSITION))
(define-quantity-direction CLOCKWISE (DIAL ANGLE))
(define-quantity-direction UP (CARRIAGE POSITION))
(define-quantity-direction SCALAR (COILS TEMPERATURE))
(define-quantity-direction SCALAR (BREAD APPEARANCE))
(define-quantity-direction SCALAR (OUTLET CHARGE))
(define-quantity-direction DOWN (EARTH GRAVITY))

 ;;;TIMELINE;;;

;t0
(start 0)
(assert-quantity-m LEVER POSITION Amount Up)
(assert-quantity-m LEVER POSITION Rate Zero)
(assert-quantity-m DIAL ANGLE Amount LM)
```

```
(assert-quantity-m DIAL ANGLE Rate Zero)
(assert-quantity-m CARRIAGE POSITION Amount Up)
(assert-quantity-m CARRIAGE POSITION Rate Zero)
(assert-quantity-m COILS TEMPERATURE Amount Off)
(assert-quantity-m COILS TEMPERATURE Rate Zero)
(assert-quantity-m BREAD APPEARANCE Amount Untoasted)
(assert-quantity-m BREAD APPEARANCE Rate Zero)
(assert-quantity-m OUTLET CHARGE Amount On)
(declare-known-cause-event
  (assert-quantity-m OUTLET CHARGE Rate Positive))
(declare-known-cause-event
  (assert-quantity-m EARTH GRAVITY Amount G))
(assert-quantity-m EARTH GRAVITY Rate Zero)

:t1
(tick 60)
(declare-known-cause-event
  (assert-quantity-m LEVER POSITION Rate Negative))
(assert-quantity-m CARRIAGE POSITION Rate Negative)

;t2
(tick 61)
(assert-quantity-m LEVER POSITION Amount Down)
(assert-quantity-m LEVER POSITION Rate Zero)
(assert-quantity-m CARRIAGE POSITION Amount Down)
(assert-quantity-m CARRIAGE POSITION Rate Zero)
(assert-quantity-m COILS TEMPERATURE Rate Positive)

;t3
(tick 66)
(declare-known-effect-event
  (assert-quantity-m BREAD APPEARANCE Rate Positive))

;t4
(tick 186)
(declare-known-effect-event
  (assert-quantity-m LEVER POSITION Rate Positive))
(assert-quantity-m CARRIAGE POSITION Rate Positive)
(assert-quantity-m COILS TEMPERATURE Amount Hot)
(assert-quantity-m COILS TEMPERATURE Rate Zero)
(assert-quantity-m BREAD APPEARANCE Amount Golden)
(assert-quantity-m BREAD APPEARANCE Rate Zero)

:t5
```

(tick 187)
(assert-quantity-m LEVER POSITION Amount Up)
(assert-quantity-m LEVER POSITION Rate Zero)
(assert-quantity-m CARRIAGE POSITION Amount Up)
(assert-quantity-m CARRIAGE POSITION Rate Zero)
(assert-quantity-m COILS TEMPERATURE Rate Negative)

;t6
(tick 450)
(assert-quantity-m COILS TEMPERATURE Amount Off)
(assert-quantity-m COILS TEMPERATURE Rate Zero)

```
;;;;;THE TIRE GAUGE;;;;;

 ;;;PHYSICAL OBJECTS;;;
(define-physical-objects SLIDE PISTON TIRE CYLINDER EARTH)

 ;;;QUANTITIES;;;
(define-quantities
  (SLIDE POSITION)
  (TIRE AMOUNT-OF-GAS)
  (EARTH GRAVITY))

 ;;;QUANTITY SPACES;;;
(define-quantity-space SLIDE POSITION Amount
  (G0 0.0) (G20 0.02) (G24 0.024) (G28 0.028)
  (G32 0.032) (G36 0.036)  (G40 0.04))
(define-quantity-space SLIDE POSITION Rate
  (NEGATIVE -0.05) (ZERO 0.0) (POSITIVE 0.05) (FAST 0.1))

(define-quantity-space TIRE AMOUNT-OF-GAS Amount
  (P0 0.0) (P2 2.0) (P4 4.0) (P6 6.0) (P8 8.0) (P10 10.0)
  (P12 12.0) (P14 14.0) (P16 16.0) (P18 18.0) (P20 20.0)
  (P22 22.0) (P24 24.0) (P26 26.0) (P28 28.0) (P30 30.0)
  (P32 32.0) (P34 34.0) (P36 36.0) (P38 38.0) (P40 40.0))
(define-quantity-space TIRE AMOUNT-OF-GAS Rate
  (NEGATIVE -0.1) (ZERO 0.0) (POSITIVE 0.1))

(define-quantity-space EARTH GRAVITY Amount
  (G 9.8))
(define-quantity-space EARTH GRAVITY Rate
  (ZERO 0.0))

 ;;;ZEROS;;;
(define-zero SLIDE POSITION G0)
(define-zero TIRE AMOUNT-OF-GAS P0)
(define-zero EARTH GRAVITY)

 ;;;CORRESPONDENCES;;;
;(define-correspondence
   (SLIDE POSITION G0) Greater (PISTON POSITION S40))

 ;;;DIRECTIONS;;;
(define-direction CYLINDER-AXIS)
(define-direction TIRE-STEM-AXIS)
(define-direction DOWN)

(define-quantity-direction CYLINDER-AXIS (SLIDE POSITION))
```

```
(define-quantity-direction TIRE-STEM-AXIS (TIRE AMOUNT-OF-GAS))
(define-quantity-direction DOWN (EARTH GRAVITY))

(affirm-relation CYLINDER-AXIS 'Skewed DOWN)

 ;;;TIMELINE;;;

;t0
(start 0)
;(affirm-relation CYLINDER 'Contains PISTON)
;(consider-relation PISTON 'Attached-To SLIDE)   ;;declaration of piston
;(consider-relation PISTON 'Connected-To SLIDE)   ;;inside tire gauge
;(consider-relation PISTON 'Touches SLIDE)
(assert-quantity-m SLIDE POSITION Amount G0)
(assert-quantity-m SLIDE POSITION Rate Zero)
(assert-quantity-m TIRE AMOUNT-OF-GAS Amount P28)
(assert-quantity-m TIRE AMOUNT-OF-GAS Rate Zero)
(declare-known-cause-event
  (assert-quantity-m EARTH GRAVITY Amount G))
(assert-quantity-m EARTH GRAVITY Rate Zero)

 :t1
(tick 60)
(affirm-relation TIRE 'Joined-To CYLINDER)
(affirm-relation CYLINDER-AXIS 'Opposite TIRE-STEM-AXIS)
(declare-known-cause-event
  (assert-quantity-m TIRE AMOUNT-OF-GAS Rate Negative))

 ;t2
(tick 60.1)
(declare-known-effect-event
  (assert-quantity-m SLIDE POSITION Rate Positive))

 ;t3
(tick 60.2)
(assert-quantity-m SLIDE POSITION Amount G28)
(declare-known-effect-event
  (assert-quantity-m SLIDE POSITION Rate Zero))
(assert-quantity-m TIRE AMOUNT-OF-GAS Amount P28)
(assert-quantity-m TIRE AMOUNT-OF-GAS Rate Zero)

 ;t4
(tick 70)
(deny-relation TIRE 'Joined-To CYLINDER)
```

```
(deny-relation CYLINDER-AXIS 'Opposite TIRE-STEM-AXIS)

 ;t5
(tick 120)
(declare-known-cause-event
  (assert-quantity-m SLIDE POSITION Rate Negative))

 :t6
(tick 120.1)
(assert-quantity-m SLIDE POSITION Amount G0)
(assert-quantity-m SLIDE POSITION Rate Zero)
```

```
;;;;;THE BICYCLE DRIVE:;;;;

 ;;;PHYSICAL OBJECTS:;;;
(define-physical-objects PEDAL SPROCKET HUB)

 ;;;QUANTITIES:;;;
(define-quantities
 (PEDAL ANGLE)
 (SPROCKET ANGLE)
 (HUB ANGLE))

 ;;;QUANTITY SPACES:;;;
(define-quantity-space PEDAL ANGLE Amount
 (BACK (* *pi* 0.0)) (TOP (// *pi* 2)) (FRONT *pi*)
 (BOTTOM (* (// *pi* 2) 3)) (BACK (* *pi* 0.0)))
(define-quantity-space PEDAL ANGLE Rate
 (NEGATIVE (minus (// *pi* 2))) (ZERO (* *pi* 0.0))
 (POSITIVE (// *pi* 2)))

(define-quantity-space SPROCKET ANGLE Amount
 (BACK (* *pi* 0.0)) (TOP (// *pi* 2)) (FRONT *pi*)
 (BOTTOM (* (// *pi* 2) 3)) (BACK (* *pi* 0.0)))
(define-quantity-space SPROCKET ANGLE Rate
 (NEGATIVE (minus (// *pi* 2))) (ZERO (* *pi* 0.0))
 (POSITIVE (// *pi* 2)))

(define-quantity-space HUB ANGLE Amount
 (BACK (* *pi* 0.0)) (TOP (// *pi* 2)) (FRONT *pi*)
 (BOTTOM (* (// *pi* 2) 3)) (BACK (* *pi* 0.0)))
(define-quantity-space HUB ANGLE Rate
 (NEGATIVE (minus (// *pi* 2))) (ZERO (* *pi* 0.0))
 (POSITIVE (// *pi* 2)))

 ;;;ZEROS;;;
(define-zero PEDAL ANGLE BACK)
(define-zero SPROCKET ANGLE BACK)
(define-zero HUB ANGLE BACK)

 ;;;DIRECTIONS;;;
(define-direction COUNTER-CLOCKWISE)

(define-quantity-direction COUNTER-CLOCKWISE (PEDAL ANGLE))
(define-quantity-direction COUNTER-CLOCKWISE (SPROCKET ANGLE))
(define-quantity-direction COUNTER-CLOCKWISE (HUB ANGLE))

 ;;;TIMELINE;;;
```

```
;t0
(start 0)
(affirm-relation PEDAL 'Attached-To SPROCKET)
(deny-relation PEDAL 'Connected-To SPROCKET)
(deny-relation PEDAL 'Touches SPROCKET)
(consider-relation SPROCKET 'Attached-To HUB)
(consider-relation SPROCKET 'Connected-To HUB)
(consider-relation SPROCKET 'Touches HUB)
(deny-relation PEDAL 'Attached-To HUB)
(deny-relation PEDAL 'Connected-To HUB)
(deny-relation PEDAL 'Touches HUB)
(assert-quantity-m PEDAL ANGLE Amount Top)
(assert-quantity-m PEDAL ANGLE Rate Zero)
(assert-quantity-m SPROCKET ANGLE Amount Front)
(assert-quantity-m SPROCKET ANGLE Rate Zero)
(assert-quantity-m HUB ANGLE Amount Back)
(assert-quantity-m HUB ANGLE Rate Zero)

;t1
(tick 60)
(declare-known-cause-event
  (assert-quantity-m PEDAL ANGLE Rate Positive))
(assert-quantity-m SPROCKET ANGLE Rate Positive)

;t2
(tick 61)
(declare-known-effect-event
  (assert-quantity-m HUB ANGLE Rate Positive))

;t3
(tick 70)
(assert-quantity-m PEDAL ANGLE Amount Front)
(declare-known-cause-event
  (assert-quantity-m PEDAL ANGLE Rate Zero))
(assert-quantity-m SPROCKET ANGLE Amount Bottom)
(assert-quantity-m SPROCKET ANGLE Rate Zero)

;t4
(tick 80)
(declare-known-cause-event
  (assert-quantity-m PEDAL ANGLE Rate Negative))
(assert-quantity-m SPROCKET ANGLE Rate Negative)

;t5
```

```
(tick 81)
(assert-quantity-m PEDAL ANGLE Amount Top)
(declare-known-cause-event
  (assert-quantity-m PEDAL ANGLE Rate Zero))
(assert-quantity-m SPROCKET ANGLE Amount Front)
(assert-quantity-m SPROCKET ANGLE Rate Zero)
(assert-quantity-m HUB ANGLE Amount Bottom)
(declare-known-effect-event
  (assert-quantity-m HUB ANGLE Rate Zero))
```

;;;;**THE REFRIGERATOR**;;;;;

;;;**PHYSICAL OBJECTS**;;;
(define-physical-objects INTERIOR EXTERIOR OUTLET)

;;;**QUANTITIES**;;;
(define-quantities
 (INTERIOR TEMPERATURE)
 (EXTERIOR TEMPERATURE)
 (OUTLET CHARGE))

;;;**QUANTITY SPACES**;;;
(define-quantity-space INTERIOR TEMPERATURE Amount
 (FROZEN -5.0 5.0) (COLD 5.0 20.0) (AMBIENT 20.0 25.0))
(define-quantity-space INTERIOR TEMPERATURE Rate
 (NEGATIVE -1.0) (ZERO 0.0) (POSITIVE 1.0))

(define-quantity-space EXTERIOR TEMPERATURE Amount
 (AMBIENT 20.0 25.0) (WARM 25.0 40.0) (HOT 40.0 100.0))
(define-quantity-space EXTERIOR TEMPERATURE Rate
 (NEGATIVE -1.0) (ZERO 0.0) (POSITIVE 1.0))

(define-quantity-space OUTLET CHARGE Amount
 (ON 100.0))
(define-quantity-space OUTLET CHARGE Rate
 (ZERO 0.0) (POSITIVE 100.0))

;;;**ZEROS**;;;
(define-zero INTERIOR TEMPERATURE FROZEN)
(define-zero EXTERIOR TEMPERATURE)
(define-zero OUTLET CHARGE)

;;;**DIRECTIONS**;;;
(define-direction SCALAR)

(define-quantity-direction SCALAR (INTERIOR TEMPERATURE))
(define-quantity-direction SCALAR (EXTERIOR TEMPERATURE))
(define-quantity-direction SCALAR (OUTLET CHARGE))

;;;**TIMELINE**;;;

;t0
(start 0)
(assert-quantity-m INTERIOR TEMPERATURE Amount Cold)
(assert-quantity-m INTERIOR TEMPERATURE Rate Zero)
(assert-quantity-m EXTERIOR TEMPERATURE Amount Ambient)

```
(assert-quantity-m EXTERIOR TEMPERATURE Rate Zero)
(declare-known-cause-event
  (assert-quantity-m OUTLET CHARGE Amount On))
(assert-quantity-m OUTLET CHARGE Rate Positive)

 ;t1
(tick 1)
(assert-quantity-m INTERIOR TEMPERATURE Rate Positive)

 ;t2
(tick 60)
(declare-known-effect-event
  (assert-quantity-m INTERIOR TEMPERATURE Rate Negative))

 ;t3
(tick 61)
(declare-known-effect-event
  (assert-quantity-m EXTERIOR TEMPERATURE Rate Positive))
```

```
;;;;;THE HOME HEATING SYSTEM;;;;;

 ;;;PHYSICAL OBJECTS;;;
(define-physical-objects FURNACE RADIATOR ROOM OUTLET EARTH)

 ;;;QUANTITIES;;;
(define-quantities
 (FURNACE TEMPERATURE)
 (RADIATOR TEMPERATURE)
 (ROOM TEMPERATURE)
 (OUTLET CHARGE)
 (EARTH GRAVITY))

 ;;;QUANTITY SPACES;;;
(define-quantity-space FURNACE TEMPERATURE Amount
 (OFF 10.0 25.0) (ON 80.0 90.0))
(define-quantity-space FURNACE TEMPERATURE Rate
 (NEGATIVE -1.0) (ZERO 0.0) (POSITIVE 1.0))

(define-quantity-space RADIATOR TEMPERATURE Amount
 (COLD 15.0 25.0) (WARM 25.0 80.0) (HOT 80.0 90.0))
(define-quantity-space RADIATOR TEMPERATURE Rate
 (NEGATIVE -1.0) (ZERO 0.0) (POSITIVE 1.0))

(define-quantity-space ROOM TEMPERATURE Amount
 (COOL 15.0 20.0) (NICE 20.0) (WARM 20.0 25.0))
(define-quantity-space ROOM TEMPERATURE Rate
 (NEGATIVE -0.0001) (ZERO 0.0) (POSITIVE 0.01))

(define-quantity-space OUTLET CHARGE Amount
 (ON 100.0))
(define-quantity-space OUTLET CHARGE Rate
 (ZERO 0.0) (POSITIVE 100.0))

(define-quantity-space EARTH GRAVITY Amount
 (G 9.8))
(define-quantity-space EARTH GRAVITY Rate
 (ZERO 0.0))

 ;;;ZEROS;;;
(define-zero FURNACE TEMPERATURE)
(define-zero RADIATOR TEMPERATURE)
(define-zero ROOM TEMPERATURE)
(define-zero OUTLET CHARGE)
(define-zero EARTH GRAVITY)
```

```
;;;DIRECTIONS;;;
(define-direction SCALAR)
(define-direction DOWN)

(define-quantity-direction SCALAR (FURNACE TEMPERATURE))
(define-quantity-direction SCALAR (RADIATOR TEMPERATURE))
(define-quantity-direction SCALAR (ROOM TEMPERATURE))
(define-quantity-direction SCALAR (OUTLET CHARGE))
(define-quantity-direction DOWN (EARTH GRAVITY))

;;;TIMELINE;;;

;t0
(start 0)
(consider-relation FURNACE 'Joined-To RADIATOR)
(consider-relation FURNACE 'Connected-To RADIATOR)
(deny-relation FURNACE 'Line-of-Sight-To RADIATOR)
(deny-relation FURNACE 'Joined-To ROOM)
(deny-relation FURNACE 'Connected-To ROOM)
(deny-relation FURNACE 'Line-of-Sight-To ROOM)
(deny-relation RADIATOR 'Joined-To ROOM)
(consider-relation RADIATOR 'Connected-To ROOM)
(consider-relation RADIATOR 'Line-of-Sight-To ROOM)
(assert-quantity-m FURNACE TEMPERATURE Amount Off)
(assert-quantity-m FURNACE TEMPERATURE Rate Zero)
(assert-quantity-m RADIATOR TEMPERATURE Amount Cold)
(assert-quantity-m RADIATOR TEMPERATURE Rate Zero)
(assert-quantity-m ROOM TEMPERATURE Amount Nice)
(assert-quantity-m ROOM TEMPERATURE Rate Zero)
(declare-known-cause-event
  (assert-quantity-m OUTLET CHARGE Amount On))
(assert-quantity-m OUTLET CHARGE Rate Positive)
(declare-known-cause-event
  (assert-quantity-m EARTH GRAVITY Amount G))
(assert-quantity-m EARTH GRAVITY Rate Zero)

;t1
(tick 1)
(assert-quantity-m ROOM TEMPERATURE Rate Negative)

;t2
(tick 21600)
(assert-quantity-m ROOM TEMPERATURE Amount Cool)
```

```
;t3
(tick 21660)
(assert-quantity-m FURNACE TEMPERATURE Rate Positive)

;t4
(tick 21780)
(assert-quantity-m FURNACE TEMPERATURE Amount On)
(assert-quantity-m FURNACE TEMPERATURE Rate Zero)

;t5
(tick 21960)
(assert-quantity-m RADIATOR TEMPERATURE Rate Positive)

;t6
(tick 21970)
(declare-known-effect-event
 (assert-quantity-m ROOM TEMPERATURE Rate Positive))

;t7
(tick 22140)
(assert-quantity-m RADIATOR TEMPERATURE Amount Hot)
(assert-quantity-m RADIATOR TEMPERATURE Rate Zero)

;t8
(tick 22570)
(assert-quantity-m ROOM TEMPERATURE Amount Nice)
(assert-quantity-m ROOM TEMPERATURE Rate Zero)
```

Appendix B: The Vocabulary of Mechanisms

```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;CAUSAL MECHANISMS;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

 ;;;;;;;;;;PROPAGATION;;;;;;;;;;
(DefMechanism PROPAGATION Causal
 ())

 ;;;;;;;;;;TRANSFORMATION;;;;;;;;;;
(DefMechanism TRANSFORMATION Causal
 ())

;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;PROPAGATIONS;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

 ;;;MECHANICAL COUPLING;;;
(DefMechanism MECHANICAL-COUPLING Causal
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range 0.1 *c*)
  :sign 'Positive
  :deflection 'Parallel
  :efficiency (make-range 1.0)
  :alignment '(Less Greater)
  :bias '(Up-Up Down-Down)
  :medium '(Attached-To Connected-To Touches))
 (Components
   Propagation))

 ;;;RIGID COUPLING;;;
(DefMechanism RIGID-COUPLING Causal
 (
  :time-constant (make-range *c*)
  :medium 'Attached-To)
 (Components
   Mechanical-Coupling))
```

```
;;;NON-RIGID COUPLING;;;
(DefMechanism NON-RIGID-COUPLING Causal
 (
  :time-constant (make-range 0.1 *c*)
  :alignment 'Greater
  :medium 'Connected-To)
 (Components
   Mechanical-Coupling))


;;;CONTACT COUPLING;;;
(DefMechanism CONTACT-COUPLING Causal
 (
  :time-constant (make-range 0.1 *c*)
  :alignment 'Less
  :medium 'Touches)
 (Components
   Mechanical-Coupling))


;;;FORWARD RATCHET;;;
(DefMechanism FORWARD-RATCHET Causal
 (
  :time-constant (make-range *c*)
  :bias 'Up-Up
  :medium 'Attached-To)
 (Components
   Mechanical-Coupling))


;;;BACKWARD RATCHET;;;
(DefMechanism BACKWARD-RATCHET Causal
 (
  :time-constant (make-range *c*)
  :bias 'Down-Down
  :medium 'Attached-To)
 (Components
   Mechanical-Coupling))
```

```
;;;ROTARY COUPLING:::
(DefMechanism ROTARY-COUPLING Causal
 (
  :independent-quantity-type 'Angle
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Angle
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range 0.1 *c*)
  :sign 'Positive
  :deflection 'Parallel
  :efficiency (make-range 0.01 100.0)
  :alignment '(Less Greater)
  :bias '(Up-Up Down-Down)
  :medium '(Attached-To Connected-To Touches))
 (Components
  Propagation))
;;;RIGID ROTARY COUPLING:::
(DefMechanism RIGID-ROTARY-COUPLING Causal
 (
  :time-constant (make-range *c*)
  :medium 'Attached-To)
 (Components
  Rotary-Coupling))
;;;NON-RIGID ROTARY COUPLING:::
(DefMechanism NON-RIGID-ROTARY-COUPLING Causal
 (
  :time-constant (make-range 0.1 *c*)
  :alignment 'Greater
  :medium 'Connected-To)
 (Components
  Rotary-Coupling))
;;;CONTACT ROTARY COUPLING:::
(DefMechanism CONTACT-ROTARY-COUPLING Causal
 (
  :time-constant (make-range 0.1 *c*)
  :alignment 'Less
  :medium 'Touches)
 (Components
  Rotary-Coupling))
```

```
;;;FORWARD ROTARY RATCHET:::
(DefMechanism FORWARD-ROTARY-RATCHET Causal
 (
  :time-constant (make-range *c*)
  :bias 'Up-Up
  :medium 'Attached-To)
 (Components
   Rotary-Coupling))


;;;BACKWARD ROTARY RATCHET:::
(DefMechanism BACKWARD-ROTARY-RATCHET Causal
 (
  :time-constant (make-range *c*)
  :bias 'Down-Down
  :medium 'Attached-To)
 (Components
   Rotary-Coupling))


;;;ELECTRICITY:::
(DefMechanism ELECTRICITY Causal
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Charge
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)  :(depends on resistors, transformers)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Connected-To)
 (Components
   Propagation))
```

```
;;;HEAT TRANSFER:::
(DefMechanism HEAT-TRANSFER Causal
 (
  :independent-quantity-type 'Temperature
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range 0.01 *c*)
  :sign '(Negative Positive)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.001 0.5)
  :alignment 'Greater
  :bias '(Down-Down Down-Up Up-Down Up-Up)
  :medium '(Connected-To Line-of-Sight-To))
 (Components
  Propagation))


;;;CONDUCTIVE HEAT EXCHANGE:::
(DefMechanism CONDUCTIVE-HEAT-EXCHANGE Causal
 (
  :time-constant (make-range 0.01 0.1)
  :sign 'Negative
  :efficiency (make-range 0.01 0.5)  ;(depends on thermal conductivity)
  :bias '(Down-Up Up-Down)
  :medium 'Connected-To)
 (Components
  Heat-Transfer))


;;;CONDUCTIVE HEAT FLOW:::
(DefMechanism CONDUCTIVE-HEAT-FLOW Causal
 (
  :time-constant (make-range 0.01 0.1)
  :sign 'Positive
  :efficiency (make-range 0.01 0.5)  ;(depends on thermal conductivity)
  :bias '(Down-Down Up-Up)
  :medium 'Connected-To)
 (Components
  Heat-Transfer))
```

```
;;;RADIATIVE HEAT EXCHANGE;;;
(DefMechanism RADIATIVE-HEAT-EXCHANGE Causal
 (
  :time-constant (make-range 0.01 *c*)
  :sign 'Negative
  :efficiency (make-range 0.001 0.1)  ;(inverse square)
  :bias '(Down-Up Up-Down)
  :medium 'Line-of-Sight-To)
 (Components
   Heat-Transfer))


;;;RADIATIVE HEAT FLOW;;;
(DefMechanism RADIATIVE-HEAT-FLOW Causal
 (
  :time-constant (make-range 0.01 *c*)
  :sign 'Positive
  :efficiency (make-range 0.001 0.1)  ;(inverse square)
  :bias '(Down-Down Up-Up)
  :medium 'Line-of-Sight-To)
 (Components
   Heat-Transfer))


;;;LIGHT TRANSMISSION;;;
(DefMechanism LIGHT-TRANSMISSION Causal
 (
  :independent-quantity-type 'Intensity
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Intensity
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.001 0.1)  ;(inverse square)
  :alignment 'Greater
  :bias '(Down-Down Up-Up)
  :medium 'Line-of-Sight-To)
 (Components
   Propagation))
```

```
;;;GAS TRANSFER;;;
(DefMechanism GAS-TRANSFER Causal
 (
  :independent-quantity-type 'Amount-of-Gas
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Amount-of-Gas
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range 0.1 100.0)
  :sign '(Negative Positive)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 1.0)  ;(gas is compressible)
  :alignment '(Less Greater)
  :bias '(Down-Down Down-Up Up-Down Up-Up)
  :medium 'Joined-To)
 (Components
  Propagation))
```

```
;;;GAS EXCHANGE;;;
(DefMechanism GAS-EXCHANGE Causal
 (
  :sign 'Negative
  :bias '(Down-Up Up-Down))
 (Components
  Gas-Transfer))
```

```
;;;GAS FLOW;;;
(DefMechanism GAS-FLOW Causal
 (
  :sign 'Positive
  :bias '(Down-Down Up-Up))
 (Components
  Gas-Transfer))
```

```
 ;;;FLUID TRANSFER;;;
(DefMechanism FLUID-TRANSFER Causal
 (
  :independent-quantity-type 'Amount-of-Fluid
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Amount-of-Fluid
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range 0.1 100.0)
  :sign '(Negative Positive)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 1.0)  ;(fluids are not compressible)
  :alignment '(Less Greater)
  :bias '(Down-Down Down-Up Up-Down Up-Up)
  :medium 'Joined-To)
 (Components
  Propagation))


 ;;;FLUID EXCHANGE;;;
(DefMechanism FLUID-EXCHANGE Causal
 (
  :sign 'Negative
  :bias '(Down-Up Up-Down))
 (Components
  Fluid-Transfer))


 ;;;FLUID FLOW;;;
(DefMechanism FLUID-FLOW Causal
 (
  :sign 'Positive
  :bias '(Down-Down Up-Up))
 (Components
  Fluid-Transfer))


;;;
```

```
;;;ELECTRO-MECHANICAL:;;
(DefMechanism ELECTRO-MECHANICAL Causal
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)  :(depends on transmission ratios)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;ELECTRO-ROTARY:;;
(DefMechanism ELECTRO-ROTARY Causal
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Angle
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)  :(depends on transmission ratios)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;ELECTRO-PHOTIC;;;
(DefMechanism ELECTRO-PHOTIC Causal
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Intensity
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)  ;(depends on resistivity)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;ELECTRO-THERMAL;;;
(DefMechanism ELECTRO-THERMAL Causal
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)  ;(depends on resistivity)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;PHOTO-CHEMICAL;;;
(DefMechanism PHOTO-CHEMICAL Causal
 (
  :independent-quantity-type 'Intensity
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Appearance
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.001 0.1)  ;(depends on ASA rating)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;THERMO-CHEMICAL;;;
(DefMechanism THERMO-CHEMICAL Causal
 (
  :independent-quantity-type 'Temperature
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Appearance
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.1 10.0)  ;(depends on chemistry)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Same)
 (Components
  Transformation)
```

```
;;;MECHANICAL EXPANSION;;;
(DefMechanism MECHANICAL-EXPANSION Causal
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Pressure
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign '(Negative Positive)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)
  :alignment '(Less Equal Greater)
  :bias '(Down-Down Down-Up Up-Down Up-Up)
  :medium 'Same)
 (Components
  Transformation))
```

```
;;;EXPANSION;;;
(DefMechanism EXPANSION Causal
 (
  :bias '(Down-Down Up-Down))
 (Components
  Mechanical-Expansion))
```

```
;;;COMPRESSION;;;
(DefMechanism COMPRESSION Causal
 (
  :bias '(Down-Up Up-Up))
 (Components
  Mechanical-Expansion))
```

```
;;;THERMAL COMPRESSION:::
(DefMechanism THERMAL-COMPRESSION Causal
 (
  :independent-quantity-type 'Temperature
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Pressure
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.01 100.0)
  :alignment '(Less Equal Greater)
  :bias '(Down-Down Up-Up)
  :medium 'Same)
 (Components
  Transformation))
;;;THERMAL EXPANSION:;;
(DefMechanism THERMAL-EXPANSION Causal
 (
  :independent-quantity-type 'Temperature
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range 0.00001 0.001)  ;(depends on expansion coefficient)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Same)
 (Components
  Transformation))

;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;FORCES:::::::::::::::::::::::::::::
 ;;;;;;;;;;FORCE;;;;;;;;;;
(DefMechanism FORCE Causal
 ()
 (Components
  Propagation))
```

```
;;;GRAVITY;;;
(DefMechanism GRAVITY Causal
 (
  :independent-quantity-type 'Gravity
  :independent-quantity-order 'Amount
  :independent-quantity-space '((G 9.8))
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection 'Parallel
  :efficiency (make-range 0.5 10.0)  :(approximates acceleration)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Reaches)
 (Components
  Force))


;;;SPRING;;;
(DefMechanism SPRING Causal
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((UNLOADED *zero*) (LOADED *zero* 0.1))
  :independent-quantity-zero 'UNLOADED
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Negative
  :deflection 'Opposite
  :efficiency (make-range 1.0 100.0)  :(depends on spring constant)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Same)
 (Components
  Force))
```

```
;;;PNEUMATIC;;;
(DefMechanism PNEUMATIC Causal
 (
  :independent-quantity-type 'Amount-of-Gas
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection 'Parallel
  :efficiency (make-range 0.1 10.0)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Contains)
 (Components
  Force))

;;;HYDRAULIC;;;
(DefMechanism HYDRAULIC Causal
 (
  :independent-quantity-type 'Amount-of-Fluid
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Positive
  :deflection 'Parallel
  :efficiency (make-range 0.1 10.0)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Contains)
 (Components
  Force))

;;;
```

```
;;;SWITCH;;;
(DefMechanism SWITCH Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero* 0.01) (OPEN 0.01))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Charge
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero  ;(cannot explain non-zero effect alone)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)  :(cannot explain non-zero effect alone)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Connected-To)
 (Components
  Enablement))

;;;LATCH;;;
(DefMechanism LATCH Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* 0.1))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Position
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Attached-To)
 (Components
  Enablement))
```

```
;;;ROTARY LATCH;;;
(DefMechanism ROTARY-LATCH Enablement
 (
  :independent-quantity-type 'Angle
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* (// *pi* 2)))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Angle
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Attached-To)
 (Components
  Enablement))

;;;VENT;;;
(DefMechanism VENT Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* 0.1))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Spans)
 (Components
  Enablement))
```

```
;;;SHUTTER;;;
(DefMechanism SHUTTER Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* 0.01))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Intensity
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias 'Up-Up
  :medium 'Spans)
 (Components
  Enablement))

;;;PNEUMATIC VALVE;;;
(DefMechanism PNEUMATIC-VALVE Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* 0.1))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Amount-of-Gas
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Spans)
 (Components
  Enablement))
```

```
;;;HYDRAULIC VALVE;;;
(DefMechanism HYDRAULIC-VALVE Enablement
 (
  :independent-quantity-type 'Position
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((CLOSED *zero*) (OPEN *zero* 0.1))
  :independent-quantity-zero 'CLOSED
  :dependent-quantity-type 'Amount-of-Fluid
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Spans)
 (Components
  Enablement))


;;;PHASE CHANGE;;;
(DefMechanism PHASE-CHANGE Enablement
 (
  :independent-quantity-type 'Pressure
  :independent-quantity-order 'Amount
  :independent-quantity-space
    '((GAS *zero* 10.0) (LIQUID 10.0 100.0))
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Down-Down Up-Up)
  :medium 'Same)
 (Components
  Enablement))
```

```
;;;CONDENSATION:::
(DefMechanism CONDENSATION Enablement
 (
  :independent-quantity-zero 'GAS
  :bias 'Up-Up)
 (Components
  Phase-Change))
```

```
;;;EVAPORATION:::
(DefMechanism EVAPORATION Enablement
 (
  :independent-quantity-zero 'LIQUID
  :bias 'Down-Down)
 (Components
  Phase-Change))
```

```
;;;FAN;;;
(DefMechanism FAN Enablement
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Amount-of-Gas
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Same)
 (Components
  Enablement))
```

```
;;;PUMP;;;
(DefMechanism PUMP Enablement
 (
  :independent-quantity-type 'Charge
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Amount-of-Fluid
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Down Up-Up)
  :medium 'Same)
 (Components
   Enablement))
```

```
;;;GAS FALL;;;
(DefMechanism GAS-FALL Enablement
 (
  :independent-quantity-type 'Gravity
  :independent-quantity-order 'Amount
  :independent-quantity-space '((G 9.8))
  :dependent-quantity-type 'Amount-of-Gas
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection 'Parallel
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Reaches)
 (Components
   Enablement))
```

```
;;;FLUID FALL;;;
(DefMechanism FLUID-FALL Enablement
 (
  :independent-quantity-type 'Gravity
  :independent-quantity-order 'Amount
  :independent-quantity-space '((G 9.8))
  :dependent-quantity-type 'Amount-of-Fluid
  :dependent-quantity-order 'Rate
  :distance 'Different
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection 'Parallel
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Up-Up Down-Down)
  :medium 'Reaches)
 (Components
  Enablement))
```

```
;;;GAS HEAT TRANSPORT;;;
(DefMechanism GAS-HEAT-TRANSPORT Enablement
 (
  :independent-quantity-type 'Amount-of-Gas
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Down-Up Up-Up)
  :medium 'Same)
 (Components
  Enablement))
```

```
;;;FLUID HEAT TRANSPORT;;;
(DefMechanism FLUID-HEAT-TRANSPORT Enablement
 (
  :independent-quantity-type 'Amount-of-Fluid
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Temperature
  :dependent-quantity-order 'Rate
  :distance 'Same
  :time-constant (make-range *c*)
  :sign 'Zero
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero*)
  :alignment '(Less Equal Greater)
  :bias '(Down-Up Up-Up)
  :medium 'Same)
 (Components
  Enablement))
;;;;;;;;;;INTEGRATION;;;;;;;;;;
(DefMechanism INTEGRATION Integration
 (
  :independent-quantity-type 'Quantity
  :independent-quantity-order 'Rate
  :dependent-quantity-type 'Quantity
  :dependent-quantity-order 'Amount
  :distance 'Same
  :time-constant (make-range *zero* *c*)
  :sign '(Negative Zero Positive)
  :deflection '(Parallel Opposite Perpendicular Skewed)
  :efficiency (make-range *zero* *c*)
  :alignment '(Less Equal Greater)
  :bias '(Down-Down Down-Up Up-Down Up-Up)
  :medium 'Same))
```

Appendix C: Qualitative Calculi

Seed Value	*Contribution of Mechanism*	*Propagated Value*
Negative	*Negative*	*Positive*
Negative	*Zero*	*Zero*
Negative	*Positive*	*Negative*
Zero	*Negative*	*Zero*
Zero	*Zero*	*Zero*
Zero	*Positive*	*Zero*
Positive	*Negative*	*Negative*
Positive	*Zero*	*Zero*
Positive	*Positive*	*Positive*

Table C.1. Qualitative Calculus for Sign.

Seed Value	*Contribution of Mechanism*	*Propagated Value*
Parallel	*Parallel*	*Parallel*
Parallel	*Opposite*	*Opposite*
Parallel	*Perpendicular*	*Perpendicular*
Parallel	*Skewed*	*Skewed*
Opposite	*Parallel*	*Opposite*
Opposite	*Opposite*	*Parallel*
Opposite	*Perpendicular*	*Perpendicular*
Opposite	*Skewed*	*Skewed*
Perpendicular	*Parallel*	*Perpendicular*
Perpendicular	*Opposite*	*Perpendicular*
Perpendicular	*Perpendicular*	{*Parallel Opposite Perpendicular*}
Perpendicular	*Skewed*	*Skewed*
Skewed	*Parallel*	*Skewed*
Skewed	*Opposite*	*Skewed*
Skewed	*Perpendicular*	*Skewed*
Skewed	*Skewed*	{*Parallel Opposite Perpendicular Skewed*}

Table C.2. Qualitative Calculus for Direction.

Seed Value	*Contribution of Mechanism*	*Propagated Value*
Less	*Less*	*Less*
Less	*Equal*	*Less*
Equal	*Less*	*Less*
Equal	*Equal*	*Equal*
Equal	*Greater*	*Greater*
Greater	*Equal*	*Greater*
Greater	*Greater*	*Greater*

Table C.3. Qualitative Calculus for Alignment.

Seed Value	*Contribution of Mechanism*	*Propagated Value*
Negative	*Down-Down*	*Negative*
Negative	*Down-Up*	*Positive*
Zero	*Down-Down*	*Zero*
Zero	*Down-Up*	*Zero*
Zero	*Up-Down*	*Zero*
Zero	*Up-Up*	*Zero*
Positive	*Up-Down*	*Negative*
Positive	*Up-Up*	*Positive*

Table C.4. Qualitative Calculus for Bias.

Seed Value	*Contribution of Mechanism*	*Propagated Value*
Same	*Same*	*Same*
Same	*Different*	*Different*
Different	*Same*	*Different*
Different	*Different*	{*Same Different*}

Table C.5. Qualitative Calculus for Displacement.

Sign at Cause	*Relative Orientation*	*Sign at Effect*
Negative	*Parallel*	*Negative*
Negative	*Opposite*	*Positive*
Negative	*Perpendicular*	{*Negative Zero Positive*}
Negative	*Skewed*	{*Negative Zero Positive*}
Zero	*Parallel*	*Zero*
Zero	*Opposite*	*Zero*
Zero	*Perpendicular*	*Zero*
Zero	*Skewed*	*Zero*
Positive	*Parallel*	*Positive*
Positive	*Opposite*	*Negative*
Positive	*Perpendicular*	{*Negative Zero Positive*}
Positive	*Skewed*	{*Negative Zero Positive*}

Table C.6. Qualitative calculus for sign and orientation.

One Contribution	*Other Contribution*	*Combined Value*
Negative	*Negative*	*Negative*
Negative	*Zero*	*Negative*
Negative	*Positive*	{*Negative Zero Positive*}
Zero	*Negative*	*Negative*
Zero	*Zero*	*Zero*
Zero	*Positive*	*Positive*
Positive	*Negative*	{*Negative Zero Positive*}
Positive	*Zero*	*Positive*
Positive	*Positive*	*Positive*

Table C.7. Qualitative calculus for sign addition.

One Contribution	*Other Contribution*	*Combined Value*
Parallel	*Parallel*	*Parallel*
Parallel	*Opposite*	{*Parallel Opposite*}
Parallel	*Perpendicular*	*Skewed*
Parallel	*Skewed*	{*Perpendicular Skewed*}
Opposite	*Parallel*	{*Parallel Opposite*}
Opposite	*Opposite*	*Opposite*
Opposite	*Perpendicular*	*Skewed*
Opposite	*Skewed*	{*Perpendicular Skewed*}
Perpendicular	*Parallel*	*Skewed*
Perpendicular	*Opposite*	*Skewed*
Perpendicular	*Perpendicular*	*Skewed*
Perpendicular	*Skewed*	{*Parallel Opposite Perpendicular Skewed*}
Skewed	*Parallel*	{*Perpendicular Skewed*}
Skewed	*Opposite*	{*Perpendicular Skewed*}
Skewed	*Perpendicular*	{*Parallel Opposite Perpendicular Skewed*}
Skewed	*Skewed*	{*Parallel Opposite Perpendicular Skewed*}

Table C.8. Qualitative calculus for direction addition.

Appendix D: Arithmetic for Order of Magnitude Ranges

The addition rule for orders of magnitude is:

$$b^{e_1} + b^{e_2} = b^{max(e_1,e_2)}$$

The addition rule for ranges of orders of magnitude is:

$$\{\cdot\textbf{RANGE}\cdot\ b^{l_1} : b^{h_1}\} + \{\cdot\textbf{RANGE}\cdot\ b^{l_2} : b^{h_2}\} = \{\cdot\textbf{RANGE}\cdot\ b^{max(l_1,l_2)} : b^{max(h_1,h_2)}\}$$

A graphic depiction of this addition rule appears in Figure D.1.

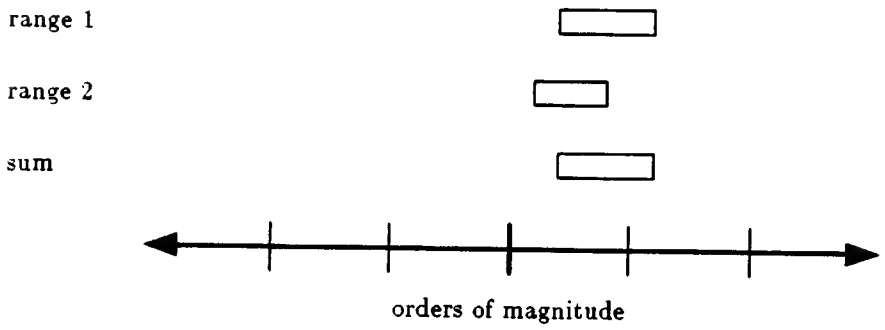


Figure D.1. Addition for ranges of orders of magnitude.

The subtraction rule for orders of magnitude is:

$$b^{e_1} - b^{e_2} =$$
if $e_1 = e_2$
then $b^{-\infty}$
else $b^{max(e_1,e_2)}$

The subtraction rule for ranges of orders of magnitude is:

$$\{\cdot\textbf{RANGE}\cdot\ b^{l_1} : b^{h_1}\} - \{\cdot\textbf{RANGE}\cdot\ b^{l_2} : b^{h_2}\} =$$
if $h_1 \geq l_2$
then $\{\cdot\textbf{RANGE}\cdot\ b^{-\infty} : b^{max(max(l_1,h_2),max(h_1,l_2))}\}$
else $\{\cdot\textbf{RANGE}\cdot\ b^{min(max(l_1,h_2),max(h_1,l_2))} : b^{max(max(l_1,h_2),max(h_1,l_2))}\}$

This subtraction rule appears in Figure D.2.

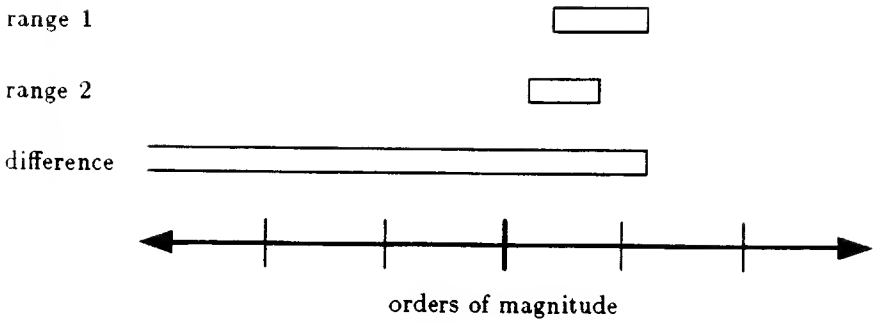


Figure D.2. Subtraction for ranges of orders of magnitude.

The multiplication rule for orders of magnitude is:

$$b^{e_1}\ b^{e_2} = b^{e_1+e_2}$$

The multiplication rule for ranges of orders of magnitude is:

$$\{\cdot\textbf{RANGE}\cdot\ b^{l_1} : b^{h_1}\}\ \{\cdot\textbf{RANGE}\cdot\ b^{l_2} : b^{h_2}\} =$$
$$\{\cdot\textbf{RANGE}\cdot\ b^{l_1+l_2} : b^{h_1+h_2}\}$$

This multiplication rule is portrayed in Figure D.3.

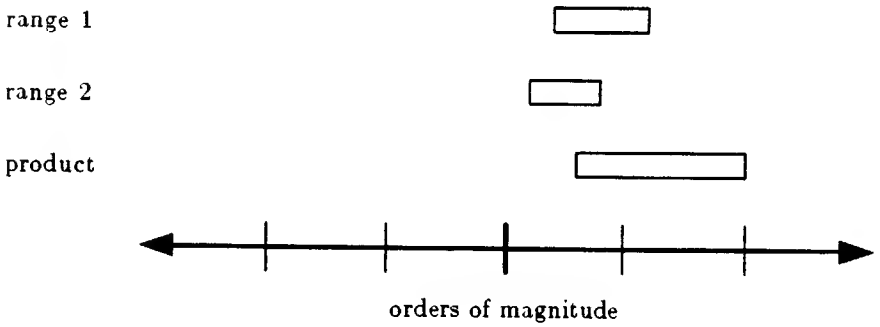


Figure D.3. Multiplication for ranges of orders of magnitude.

The division rule for orders of magnitude is:

$$b^{e_1}\ /\ b^{e_2} = b^{e_1-e_2}$$

The division rule for ranges of orders of magnitude is:

$$\{\cdot\texttt{RANGE}\cdot\ b^{l_1} : b^{h_1}\} \ / \ \{\cdot\texttt{RANGE}\cdot\ b^{l_2} : b^{h_2}\} =$$
$$\{\cdot\texttt{RANGE}\cdot\ b^{min(l_1-h_2,h_1-l_2)} : b^{max(l_1-h_2,h_1-l_2)}\}$$

This division rule is shown in Figure D.4.

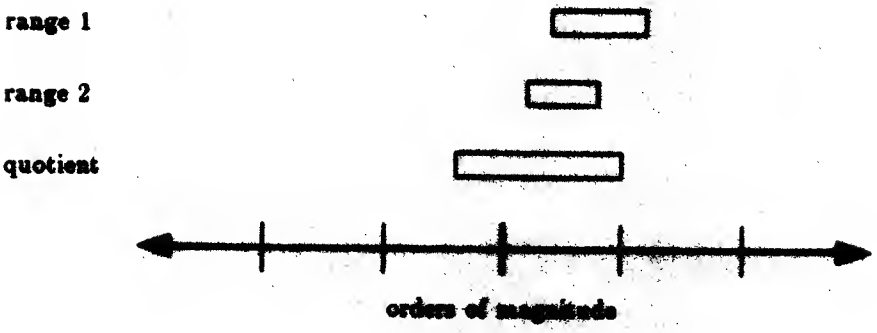


Figure D.4. Division for ranges of orders of magnitude.

Appendix E: Cause and Effect Types of Mechanisms

Cause Type	*Mechanism*	*Effect Type*
{·TYPE· *Position Rate*}	*Rigid-Coupling*	{·TYPE· *Position Rate*}
{·TYPE· *Position Rate*}	*Non-Rigid-Coupling*	{·TYPE· *Position Rate*}
{·TYPE· *Position Rate*}	*Contact-Coupling*	{·TYPE· *Position Rate*}
{·TYPE· *Position Rate*}	*Forward-Ratchet*	{·TYPE· *Position Rate*}
{·TYPE· *Position Rate*}	*Backward-Ratchet*	{·TYPE· *Position Rate*}
{·TYPE· *Angle Rate*}	*Rigid-Rotary-Coupling*	{·TYPE· *Angle Rate*}
{·TYPE· *Angle Rate*}	*Non-Rigid-Rotary-Coupling*	{·TYPE· *Angle Rate*}
{·TYPE· *Angle Rate*}	*Contact-Rotary-Coupling*	{·TYPE· *Angle Rate*}
{·TYPE· *Angle Rate*}	*Forward-Rotary-Ratchet*	{·TYPE· *Angle Rate*}
{·TYPE· *Angle Rate*}	*Backward-Rotary-Ratchet*	{·TYPE· *Angle Rate*}
{·TYPE· *Charge Rate*}	*Electricity*	{·TYPE· *Charge Rate*}
{·TYPE· *Temperature Rate*}	*Conductive-Heat-Exchange*	{·TYPE· *Temperature Rate*}
{·TYPE· *Temperature Rate*}	*Conductive-Heat-Flow*	{·TYPE· *Temperature Rate*}
{·TYPE· *Temperature Rate*}	*Radiative-Heat-Exchange*	{·TYPE· *Temperature Rate*}
{·TYPE· *Temperature Rate*}	*Radiative-Heat-Flow*	{·TYPE· *Temperature Rate*}
{·TYPE· *Intensity Rate*}	*Light-Transmission*	{·TYPE· *Intensity Rate*}
{·TYPE· *Amount-of-Gas Rate*}	*Gas-Exchange*	{·TYPE· *Amount-of-Gas Rate*}
{·TYPE· *Amount-of-Gas Rate*}	*Gas-Flow*	{·TYPE· *Amount-of-Gas Rate*}
{·TYPE· *Amount-of-Fluid Rate*}	*Fluid-Exchange*	{·TYPE· *Amount-of-Fluid Rate*}
{·TYPE· *Amount-of-Fluid Rate*}	*Fluid-Flow*	{·TYPE· *Amount-of-Fluid Rate*}
{·TYPE· *Charge Rate*}	*Electro-Mechanical*	{·TYPE· *Position Rate*}
{·TYPE· *Charge Rate*}	*Electro-Rotary*	{·TYPE· *Angle Rate*}
{·TYPE· *Charge Rate*}	*Electro-Photic*	{·TYPE· *Intensity Rate*}
{·TYPE· *Charge Rate*}	*Electro-Thermal*	{·TYPE· *Temperature Rate*}
{·TYPE· *Intensity Rate*}	*Photo-Chemical*	{·TYPE· *Appearance Rate*}
{·TYPE· *Temperature Rate*}	*Thermo-Chemical*	{·TYPE· *Appearance Rate*}
{·TYPE· *Charge Rate*}	*Expansion*	{·TYPE· *Pressure Rate*}
{·TYPE· *Charge Rate*}	*Compression*	{·TYPE· *Pressure Rate*}
{·TYPE· *Temperature Rate*}	*Thermal-Expansion*	{·TYPE· *Position Rate*}
{·TYPE· *Gravity Amount*}	*Gravity*	{·TYPE· *Position Rate*}
{·TYPE· *Position Amount*}	*Spring*	{·TYPE· *Position Rate*}
{·TYPE· *Amount-of-Gas Rate*}	*Pneumatic*	{·TYPE· *Position Rate*}
{·TYPE· *Amount-of-Fluid Rate*}	*Hydraulic*	{·TYPE· *Position Rate*}

Table E.1. Type relations for mechanisms.

Cause Type	*Mechanism*	*Effect Type*
{·TYPE· *Position Amount*}	*Switch*	{·TYPE· *Charge Rate*}
{·TYPE· *Position Amount*}	*Latch*	{·TYPE· *Position Rate*}
{·TYPE· *Angle Amount*}	*Rotary-Latch*	{·TYPE· *Angle Rate*}
{·TYPE· *Position Amount*}	*Vent*	{·TYPE· *Temperature Rate*}
{·TYPE· *Position Amount*}	*Shutter*	{·TYPE· *Intensity Rate*}
{·TYPE· *Position Amount*}	*Pneumatic-Valve*	{·TYPE· *Amount-of-Gas Rate*}
{·TYPE· *Position Amount*}	*Hydraulic-Valve*	{·TYPE· *Amount-of-Fluid Rate*}
{·TYPE· *Pressure Amount*}	*Condensation*	{·TYPE· *Heat Rate*}
{·TYPE· *Pressure Amount*}	*Evaporation*	{·TYPE· *Heat Rate*}
{·TYPE· *Charge Rate*}	*Fan*	{·TYPE· *Amount-of-Gas Rate*}
{·TYPE· *Charge Rate*}	*Pump*	{·TYPE· *Amount-of-Fluid Rate*}
{·TYPE· *Gravity Amount*}	*Gas-Fall*	{·TYPE· *Amount-of-Gas Rate*}
{·TYPE· *Gravity Amount*}	*Fluid-Fall*	{·TYPE· *Amount-of-Fluid Rate*}
{·TYPE· *Amount-of-Gas Rate*}	*Gas-Heat-Transport*	{·TYPE· *Temperature Rate*}
{·TYPE· *Amount-of-Fluid Rate*}	*Fluid-Heat-Transport*	{·TYPE· *Temperature Rate*}

Table E.1 (cont.). Type relations for enablement mechanisms.

Cause Type	*Mechanism*	*Effect Type*
{·TYPE· *Position Rate*}	*Integration*	{·TYPE· *Position Amount*}
{·TYPE· *Angle Rate*}	*Integration*	{·TYPE· *Angle Amount*}
{·TYPE· *Charge Rate*}	*Integration*	{·TYPE· *Charge Amount*}
{·TYPE· *Temperature Rate*}	*Integration*	{·TYPE· *Temperature Amount*}
{·TYPE· *Pressure Rate*}	*Integration*	{·TYPE· *Pressure Amount*}
{·TYPE· *Amount-of-Gas Rate*}	*Integration*	{·TYPE· *Amount-of-Gas Amount*}
{·TYPE· *Amount-of-Fluid Rate*}	*Integration*	{·TYPE· *Amount-of-Fluid Amount*}
{·TYPE· *Intensity Rate*}	*Integration*	{·TYPE· *Intensity Amount*}
{·TYPE· *Appearance Rate*}	*Integration*	{·TYPE· *Appearance Amount*}

Table E.1 (cont.). Type relations for temporal integration.

Appendix F: Using Causal Models in Device Monitoring

In this final appendix, I argue for the utility of causal models in a specific problem solving task concerning physical systems—the monitoring of devices. This phase of my research involved the transfer of results from my thesis to a project at the Jet Propulsion Laboratory [Doyle et al 87].

Monitoring is the detection of anomalies in the behavior of a physical system. Monitoring involves collecting measurements from sensors, combining this data into a picture of the current state of the system, and assessing any departure from nominal behavior. Traditional approaches to monitoring prove inadequate in the face of two issues: The dynamic adjustment of expectations about sensor values when the behavior of the device is too complex to enumerate beforehand, and the selective but effective interpretation of sensor readings when the number of sensors precludes comprehensive monitoring.

I explore an approach to monitoring which addressing these issues and which involves the use of causal models of devices. Model-based simulations of behavior support the dynamic adjustment of expectations about sensor values as the operating context of a device changes. Furthermore, a causal simulation which describes device events and dependencies among them supports planning decisions about how to utilize a limited numbers of sensors to verify correct operation of the device efficiently and reliably.

F.1 Motivation

Numerous on-line physical systems require round-the-clock supervision. As the complexity of devices and the number of sensors have increased, automated monitoring techniques to aid the human operator are showing signs of becoming inadequate. Both false alarms and undetected anomalies occur, increasing the burden of interpretation on the operators. As devices continue to become more complex, machines must take on a greater portion of the monitoring task if a near real-time response capability is to be maintained.

Automated monitoring becomes a matter of necessity with some physical systems. An example is the life support system for the proposed national Space Station. Clearly, the human resources onboard are too scarce to commit to the interpretation of sensors. Unfortunately, diverting this task to humans on the ground contains an element of risk. Communications do fail and telemetry is lost—and even a momentary interruption in monitoring could prove catastrophic. The only acceptable option is an onboard automated monitoring capability.

F.2 Issues

Traditional approaches to monitoring associate predefined nominal ranges with sensors. Alarms are raised whenever sensor readings fall outside these fixed ranges. This approach is appropriate for ensuring that a device does not operate outside its performance limits, but is woefully inadequate for monitoring physical systems which have multiple operating modes or which interact with their environment.

For example, consider the flow rate of coolant in a nuclear reactor. The tolerance on the safe rate of flow depends on whether or not the control rods are engaged in the reactor core. In other words, the nominal range for the relevant sensor is dynamic. A fixed range can lead to either false alarms or a potentially disastrous undetected anomaly.

In another example, consider a mobile robot traversing the surface of Mars. The monitoring system certainly should raise an alarm when the inclination of the rover approaches the point of overbalance. But in addition, the monitoring system should be able to flag even a slight tilt, albeit not immediately dangerous, when the world model indicates that the terrain is flat. The nominal sensor range for the inclinometer depends on the interaction of the rover with its environment.

Another problem in monitoring is potentially overwhelming sensor data. When the number of sensors in a physical system is numbered in the thousands, the ability to read and interpret these sensors in real-time, whether by man or machine, becomes compromised. The difficulty arises not only from bandwidth and sampling rate limitations, but also in trying to synthesize disparate sensor data into a global picture of the state of the system. Current monitoring systems do not address this problem, except through the brute force approach of faster and faster hardware—a solution which treats only the symptom.

The apparent human solution to this problem is straightforward. At any time, only a few sensors are interpreted—those which provide the most relevant data on the state of the system, in the current context. For example, when changing lanes on the freeway, a driver makes use of the side mirrors. While cruising in the fast lane, these particular sensors are sampled only infrequently, if at all.

Automated monitoring should be buttressed by a sensor planning capability in which sensors are treated as resources and context-sensitive importance criteria are used to determine which sensors to sample and interpret at any given time.

F.3 Domains

In choosing a testbed for this research we restricted our choice to domains with relevance for the Jet Propulsion Laboratory and the National Aeronautics and Space Administration. We enumerated selection criteria which reflect the issues raised by shortcomings in traditional approaches to monitoring.

Our ideal problem domain satisfies the following criteria:

- Numerous and diverse mechanisms and sensors.
- Multiple operating modes and/or interaction with environment.
- Real-time response required.
- Desirable to remove burden of interpretation from human operators.
- Comprehensive sensor interpretation difficult.

We narrowed our choice to four domains, including JPL's Space Simulator, the Deep Space Network tracking stations, the Thermal Management System of the Space Station, and the Mars Rover.

The Space Simulator is a chamber in which spacecraft and instruments can be subjected to some of the aspects of the space environment—intense cold, near vacuum, and solar radiation. The Deep Space Network is used to track and maintain communication with spacecraft. The Thermal Management System of the Space Station maintains appropriate temperatures in the areas assigned to crew, cargo, and scientific instruments. An autonomous Mars Rover offers an unprecedented challenge to monitoring technology. Given that the minimum communication time to Earth is on the order of ten minutes, there must be an onboard capability for quickly recognizing potentially dangerous situations. The conventional approach to monitoring might very well leave the rover in a state of paralysis, constantly processing false alarms.

F.4 Predictive Monitoring Based on Causal Simulation

Our approach to the monitoring problem involves the use of causal models of devices. Our claim is that causal models contain the information which provides answers to the questions "What should be happening in the device at this moment?" and "How can the correct operation of the device be verified quickly and reliably?". We term our approach *predictive monitoring.*

We envision three complementary capabilities in a predictive monitoring system: causal simulation, sensor planning, and sensor interpretation. We are developing a system which we call PREMON. See Figure F.1.

The causal simulator, given a causal model of a physical system and an initial set of events, generates predictions concerning the next expected events

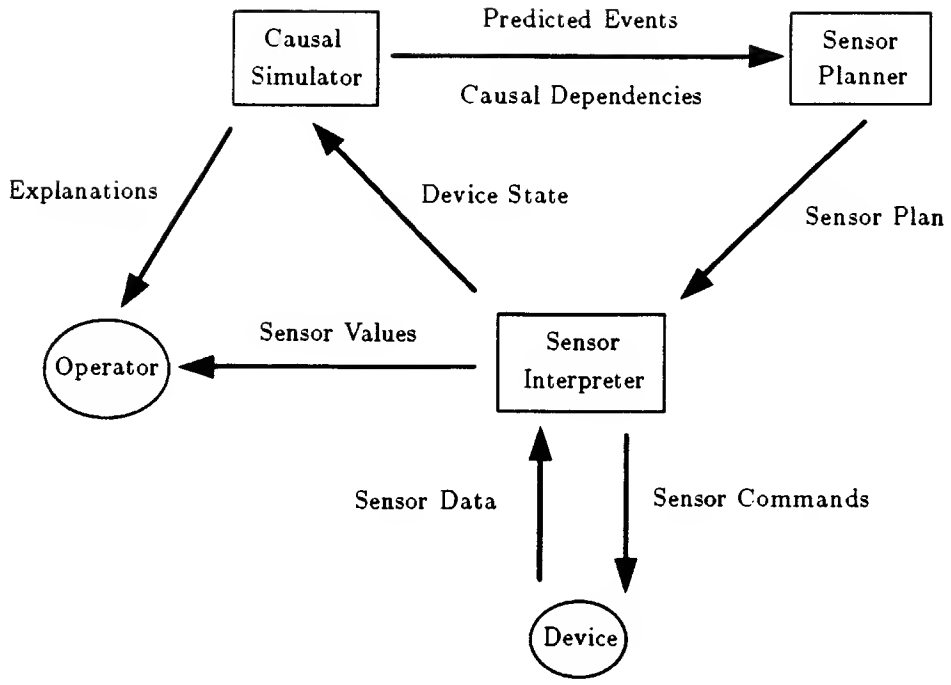


Figure F.1. A predictive monitoring system.

in the device. The causal model distinguishes different operating modes of the device and, when appropriate and possible, includes knowledge about the environment with which the physical system interacts.

The sensor planner, given this set of expectations concerning the next events in the device, makes choices about what subset of this behavior to verify, which sensors to employ, and how sensors should be sampled. These determinations are passed as a set of instructions to the sensor interpreter.

The sensor interpreter reads sensor channels as instructed by the sensor planner and compares actual sensor data with the expectations generated by the causal simulator. Discrepancies result in the raising of alarms. Finally, this most up-to-date sensor data is passed back to the causal simulator to seed the next cycle of predicting, planning, and sensing.

In the remainder of this section, I elaborate on causal simulation and sensor planning. Sensor interpretation is not treated further.

F.4.1 Causal Simulation

Simulation directly addresses the issue of changing operating contexts in monitoring. Part of the input to a simulator is the current state of the device and possibly its environment. This state specifies the operating context of the device—for example, whether the control rods are in or out of the reactor core, or whether the Mars rover is traversing terrain or collecting samples.

Predefined alarm thresholds constitute an over-summarized model of a device. They are ineffective because few sensor values can be classified *a priori* as always indicative of a problem or always indicative of correct operation. An explicit model restores the ability to evaluate sensor values in the dynamic operating context of a device. Both false alarms and undetected anomalies can be avoided, as shown in Figure F.2.

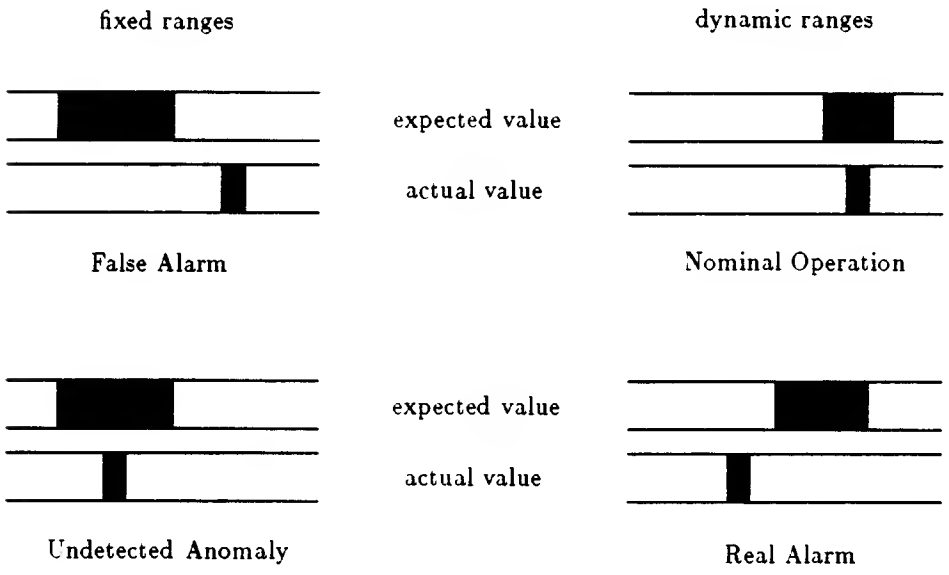


Figure F.2. Fixed vs. dynamic nominal sensor ranges.

The causal models input to the predictive monitoring system PREMON are hand-generated from the same vocabulary of mechanisms used by the program JACK. They are not generated by the causal modelling system JACK. My intent is not to validate specific causal models generated by the program JACK, but to argue

for the utility of the reasoning supported by causal models of complex real-world physical systems.

The causal simulation method used in the program PREMON is taken directly from the program JACK. Effect events are predicted from cause events by propagating values for the constraints on type, behavior, and structure across mechanisms which form the arcs of causal graphs.

Precise numerical data about the state of a device will not always be available. It is neither desirable nor likely possible to verify every aspect of a device's behavior in each predict-plan-sense cycle. The causal simulator of PREMON must be able to sustain predictions about the next events in a device in the face of spotty sensor data.

The representations for mechanisms accommodate default values for quantities. In addition, the methods for propagating qualitative regions and order of magnitude ranges allow for robust computing in the absence of numerical precision. These aspects of the causal simulation method borrowed from the program JACK enable the predictive monitoring system PREMON to generate expectations about events without acquiring precise numerical values to seed simulation; values which may be costly or impossible to come by.

F.4.2 Sensor Planning

The problem of potentially overwhelming sensor data in monitoring is addressed by introducing a sensor planning capability. Sensors are treated as information resources which need to be explicitly managed. The goal is to efficiently acquire relevant sensory information.

Our intuition is as follows: the set of sensors which provide the most direct and complete verification of the operation of a device depends, as do the values expected on those sensors, on the operating context of the device. For example, proximity detectors, tachometers, inclinometers, and accelerometers provide the most relevant sensory information for a mobile robot traversing the surface of Mars. On the other hand, when the rover is collecting samples and is stationary, force sensors, position encoders, and the vision system provide the most direct confirmation of correct operation. Just as it is unreasonable to assume that nominal sensor values can be predetermined, so it is unreasonable to assume that there is an *a priori* distinguishable subset of sensors which are sources of pertinent information for all situations.

Once again, the key to our approach is the use of a causal device model. A simulation derived from a causal model yields information about where the next changes in the values of quantities will occur. Sampling can be focused on those sensors which measure the quantities which are predicted to change.

A simulation trace also reveals causal dependencies among events in a physical system. For example, heating may lead to thermal expansion which closes a switch and turns on a motor; motion may result in the displacement of a spring which produces a restoring force and arrests the motion. Analysis of causal dependencies supports decisions about what to monitor and how carefully to monitor. The importance of events can be assessed by determining how many other events are effects or causes of a given event. In other words, the importance of an event is related to the amount of subsequent activity it supports, and the amount of activity which arranges for its occurrence. Events such as the closing of a valve and the release of a latch which lie on more than one mechanism path should be verified with care, perhaps with a battery of sensors. On the other hand, events which are side effects and do not support further activity of the device need be given only cursory attention, if at all. See Figure F.3.

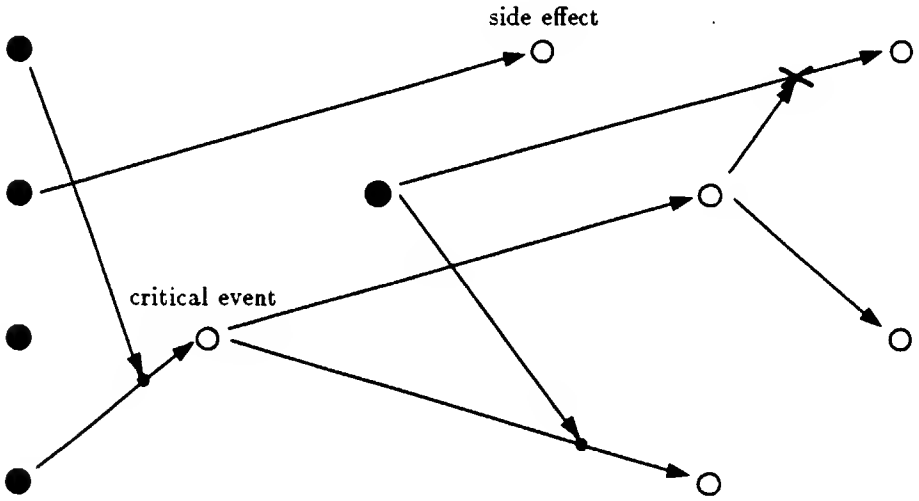


Figure F.3. Assessing the importance of events.

Our approach to sensor planning is similar to the minimum entropy method of de Kleer and Williams [de Kleer and Williams 87]. Their technique determines the best sites for test measurements in diagnosis by propagating known quantity values and component failure probabilities along causal dependencies in circuits.

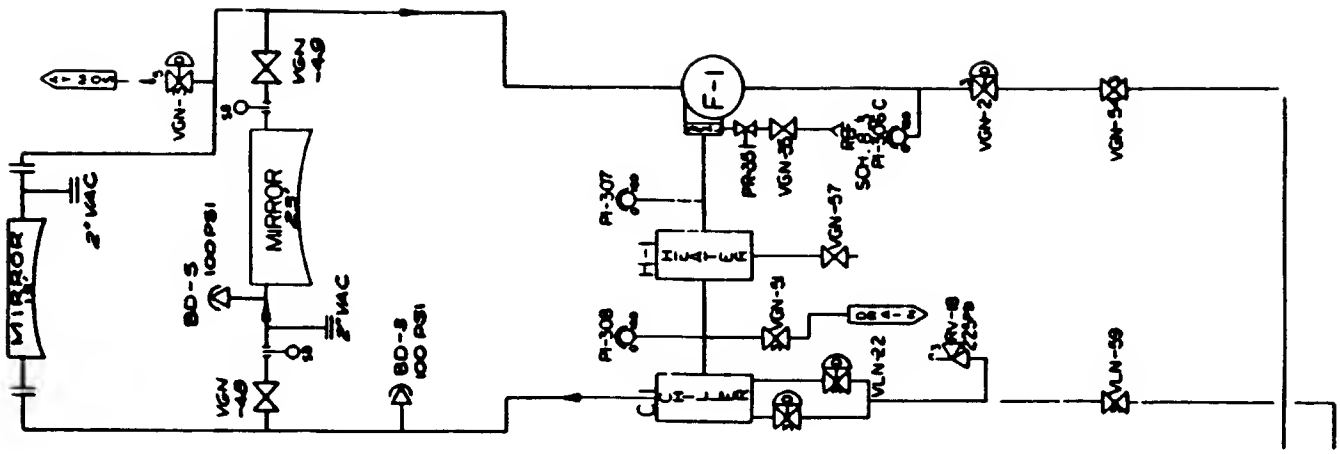


Figure F.4. Mirror cooling circuit of the JPL Space Simulator.

Sensor planning has received attention in the robot planning literature [Sacerdoti 77, Fox et al 84, Miller 85, Gini et al 85]. The roots of our own work are in a project on the monitoring of robot plan execution [Doyle et al 86]. In that project, we implemented a sensor planner called **GRIPE** which analyzes a robot task plan, inserting appropriate perception requests into the plan and generating expectations about sensor values. The program **GRIPE** has been tested successfully on a task plan which reproduces the actions taken by Space Shuttle astronauts to repair the Solar Max satellite.

Sensor planning issues have equal relevance for physical system monitoring and robot plan monitoring. We have a reasonable expectation that results from our current research effort will transfer over to the task of monitoring the execution of robot plans.

F.5 An Example: The JPL Space Simulator

In the JPL Space Simulator, a mirror is used to direct simulated solar radiation onto the spacecraft or instrument inside the chamber. This mirror must be cooled close to the temperature of the shroud which surrounds the chamber. Cold gaseous nitrogen is used as the cooling medium and is circulated by a fan. Chilling is achieved by spraying liquid nitrogen into the circulating gaseous nitrogen. Any required warming is achieved electrically. A schematic diagram of this subsystem of the JPL Space Simulator is shown in Figure F.4.

A causal simulation of the mirror cooling circuit is shown in Figure F.5. This simulation is derived from a hand-generated partial model of the circuit.

Analysis of the causal dependencies in this trace of the operation of the mirror cooling circuit identifies the flow of gas at the fan as the single most critical event in the nominal operation of the circuit. More events in the simulation trace are causes or effects of this event than of any other event. Other important events include the temperature changes which occur in the circulating gas at the chiller and at the heater.

F.6 Research Topics in Sensor Planning

In this section, I enumerate a number of issues associated with the task of monitoring physical systems.

What to Monitor?

It is neither desirable nor feasible to interpret all sensor channels of a device at all times. Choices have to be made concerning which aspects of the expected behavior of a physical system should be verified at any given time. We are developing heuristics for evaluating the relative importance of sensor data. Among these heuristics are: Changed quantity values are more important than unchanged values. Importance is related to the number of causal dependencies in which an event participates.

How to Monitor?

In many cases, the sensor most appropriate for measuring a given quantity of a physical system can be predetermined. Indeed, sensor configurations often reflect particular anticipated monitoring needs. However, for the most important events, it may be appropriate to employ multiple sensors to enhance the reliability of verification.

A different problem arises when sensors fail. The most appropriate sensor for measuring a particular quantity may be unavailable. The only recourse is to determine which sensor(s) can verify an expected event indirectly.

When to Monitor?

The issue of when and how often to monitor, is the continuous generalization of the issue of what to monitor at all. For example, the tachometers in a rover locomotion system might be monitored nearly continuously when slippery terrain is being traversed, less often on stable terrain, and very infrequently when there is no intention to move the rover.

Even a quantity which is not expected to change might be monitored at a low sampling rate, particularly if the stable value represents a state on which later events depend, or a state difficult to re-achieve.

What to Expect?

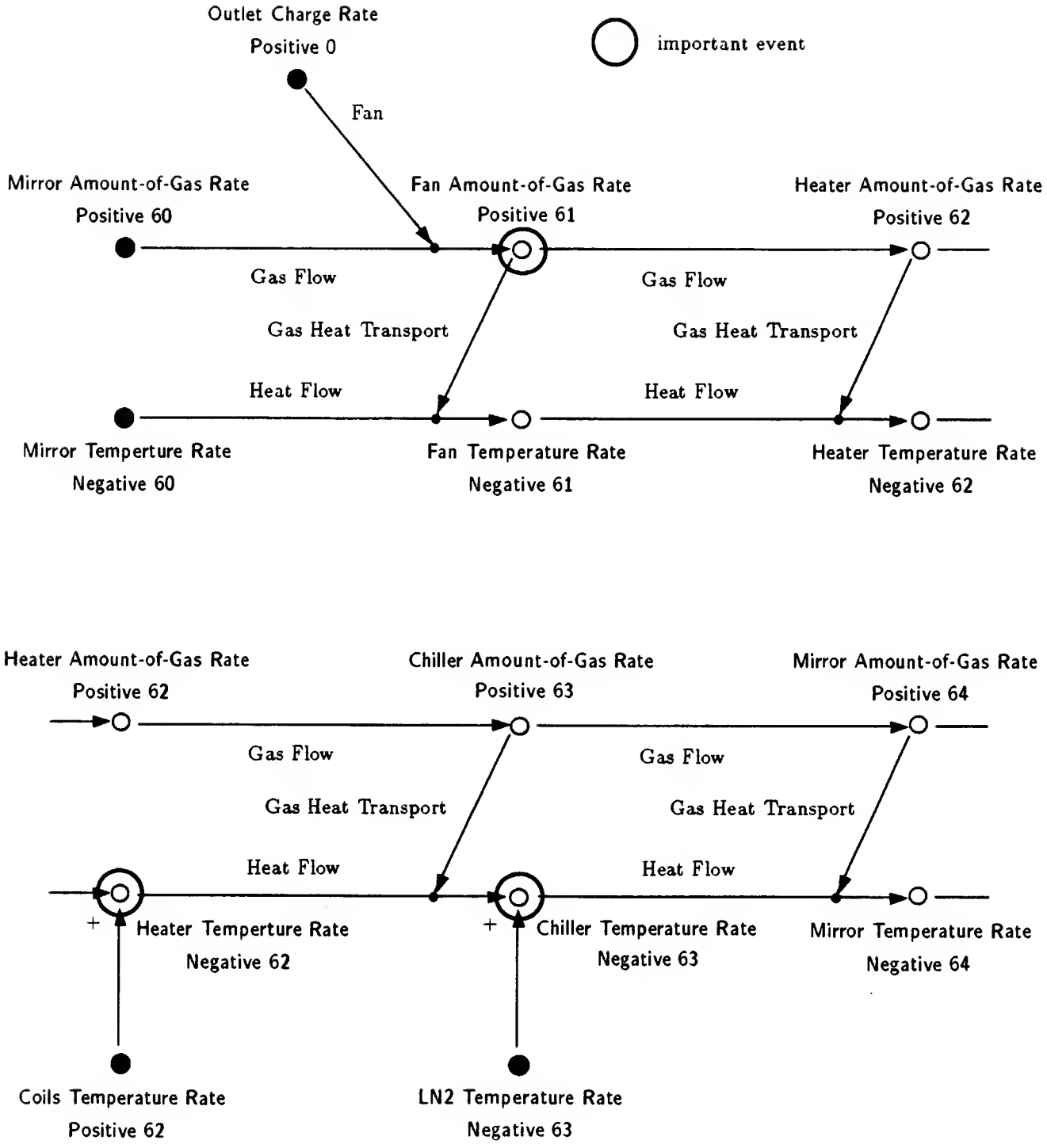


Figure F.5. Causal simulation of the mirror cooling circuit.

In most cases, expectations about sensor values can be gleaned directly from a simulation. However—particularly when sensors fail—there may be no sensor which can directly verify an anticipated event. In this case, a different event which implies the event of interest should be verified. For example, the closing of a valve might be verified indirectly by monitoring a pressure sensor upstream from the valve, should the valve's own state indicator fail. Such alternate events can be found by tracing dependencies in the simulation.

Device Operation and Sensor Planning

Decisions about how to operate a device can depend on monitoring capabilities. For example, when sensors fail, it may be prudent to operate a device in an inefficient but verifiable manner.

This blank page was inserted to preserve pagination.

CS-TR Scanning Project
Document Control Form

Date : 8/31/95

Report # AI-TR-1047

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☒ Technical Report (TR) ☐ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 213(221-images)
Not to include DOD forms, printer instructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other: Copy ?

Check each if included with document:

☒ DOD Form (2) ☐ Funding Agent Form ☒ Cover Page
☒ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages (by page number): _____

Photographs/Tonal Material (by page number): _____

Other (note description/page number):

	Description :	Page Number:
(A)	IMAGE MAP: (1) UNNUMBERED TITLE PAGE	
	(2-213) PAGES #'ED 2-213	
	(214-221) SCANCONTROL, COVER, SPINE, DOD(2), TRGT'S(3)	
(B)	MARK AT TOP OF EACH PAGE.	

Scanning Agent Signoff:

Date Received: 8/31/95 Date Scanned: 9/1/95 Date Returned: 9/7/95

Scanning Agent Signature: Michael N. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AI-TR 1047

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Hypothesizing Device Mechanisms: Opening Up the Black Box

5. TYPE OF REPORT & PERIOD COVERED: technical report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Richard James Doyle

8. CONTRACT OR GRANT NUMBER(s): N00014-85-K-0124

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Artificial Inteligence Laboratory
545 Technology Square
Cambridge, MA 02139

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

12. REPORT DATE: June 1988

13. NUMBER OF PAGES: 213

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):
Office of Naval Research
Information Systems
Arlington, VA 22217

15. SECURITY CLASS. (of this report):

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

causal reasoning
theory formation
qualitative reasoning
modelling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Abstract: Causal models of devices support many forms of problem solving in the physical system domain, such as diagnosis and monitoring. I describe an approach to forming hypotheses about hidden mechanism configurations within devices given external observations and a vocabulary of primitive mechanisms. The approach has two aspects: one involves a set of constraints drawn from physical and causal principles to prune hypotheses; the other involves an ordering on hypothesis

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 Cont.

types and a set of rules for traversing the ordering to carefully control the generation of hypotheses. The rules are all based on the principle that incomplete hypotheses exhibit characteristic deficiencies; they justify attempts to augment deficient hypotheses by extending them into more complex hypotheses.

This approach has been implemented in a causal modelling system called JACK. The program JACK generates manageably sized sets of hypotheses about the mechanisms within devices and makes fine distinctions among hypotheses. This causal modelling system reasons about the behavior of several diverse devices, constructing explanations for why a second piece of toast in a toaster comes out lighter, why the slide in a tire gauge does not slip back inside the cylinder when the gauge is removed from the tire, and how in a refrigerator a single substance can serve alternately as a heat sink for the interior and a heat source for the exterior.

I analyze the performance of the program JACK in two ways: in terms of the number of hypotheses admitted for each device example and how these hypotheses are organized in an abstraction space, and in terms of empirical results from a set of experiments which isolate the pruning power due to the different sources of constraint in my approach to the causal modelling problem. In conclusion, I show how causal models of devices produced by the program JACK can be used to support diagnosis and monitoring tasks.

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**